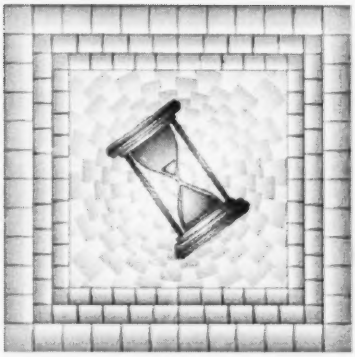
N° 62F0014MIF au catalogue — N° 17
ISSN : 1706-7731
ISBN : 0-662-76328-9

Document de recherche

Série analytique - Division des prix

Comparaisons intervilles du coût du logement au Canada : une approche hédonique



par James Chowhan et Marc Prud'homme

Division des prix
Immeuble Jean-Talon, 13e étage, Ottawa, K1A 0T6

Téléphone: 1 613 951-9606 Télécopieur: 1 613 951-1539

Toutes les opinions émises par les auteurs de ce document ne reflètent pas nécessairement celles de Statistique Canada.

Statistique Canada Statistics Canada

Canada

Comparaisons intervilles du coût du logement au Canada :

une approche hédonique

par

James Chowhan
McMaster University Research Data Centre
Bureau 217 Mills Memorial Library
McMaster University
1280 Main Street West
Hamilton (Ontario) L8S 4L6
Téléphone : 1 905 525–9140, poste 27967
Courriel : chowhan@mcmaster.ca

Marc Prud'homme
Système de comptabilité nationale
Immeuble R.H. Coats
Statistique Canada
Ottawa (Ontario)
K1A 0T6
Téléphone : 1 613 951–4415
Courriel : marc.prud'homme@statcan.ca

Publié par la Division des prix
Immeuble Jean-Talon, 13e étage, Ottawa (Ontario) K1A OT6
Télécopieur : 1 613 951-1539

N° 62F0014MIF au catalogue, série n° 17
ISSN : 1706–7731
ISBN : 0–662–76328–9

Ce document peut être consulté à l'adresse
www.statcan.ca/francais/IPS/Data/62F0014MIF2004017.htm

Avril 2004

Les opinions exprimées dans le présent document sont celles des auteurs et ne représentent pas forcément celles de la Division des prix ni celles de Statistique Canada.

Préface

La Division des prix a commencé à publier la Série analytique en décembre 1996 afin de faire part au public des travaux de recherche conceptuels et appliqués sur les indices de prix entrepris par ses employés et, parfois, par d'autres personnes au sein ou à l'extérieur de Statistique Canada.

Tous les documents sont revus par un groupe de spécialistes de Statistique Canada ou de l'extérieur. Les opinions exprimées par les auteurs ne reflètent pas forcément celles de la Division des prix ni celles de Statistique Canada.

La série analytique a pour but de diffuser les connaissances et de susciter la discussion. Les questions et commentaires concernant tout aspect de ces documents devraient être adressés au directeur, Division des prix ou au chef, Section du contrôle de la qualité et du service à la clientèle, 13^{e} étage, Immeuble Jean-Talon, Statistique Canada, Ottawa, Ontario, K1A 0T6, télécopieur : 1 613 951–1539.

Publication autorisée par le ministre responsable de Statistique Canada

© Ministre de l'Industrie, 2004

Tous droits réservés. Il est interdit de reproduire ou de transmettre le contenu de la présente publication, sous quelque forme ou par quelque moyen que ce soit, enregistrement sur support magnétique, reproduction électronique, mécanique, photographique, ou autre, ou de l'emmagasiner dans un système de recouvrement, sans l'autorisation écrite préalable des Services de concession des droits de licence, Division du marketing, Statistique Canada, Ottawa, Ontario, Canada K1A 0T6.

Résumé

Jusqu'à récemment, à cause de problèmes conceptuels, le programme des indices comparatifs des prix de détail entre les villes de la Division des prix excluait toute référence au logement. Pourtant, il est bien établi que les différences géographiques entre les niveaux de prix sont dues aux variations du coût du logement. La découverte de moyens de tenir compte du logement dans les indices fait depuis longtemps l'objet de travaux de recherche. Récemment, l'application de la modélisation hédonique à un ensemble complet de données sur les loyers a donné des résultats prometteurs qui permettent maintenant d'inclure le logement dans le programme spatial. Il est désormais possible de produire un indice de prix d'ensemble spatial qui donne une mesure nettement plus exacte des différences géographiques de prix entre les principales villes canadiennes. L'extension de l'indice d'ensemble spatial répondra sans aucun doute à la demande croissante d'information de ce genre. Une mesure plus complète des différences de prix entre les villes pourrait, par exemple, servir de déflateur pour comparer les niveaux régionaux de revenu ou pour rajuster les salaires et les traitements. Le présent document décrit la méthode utilisée pour élaborer l'élément logement du programme de l'indice spatial courant. Une approche fondée sur l'équivalence locative est adoptée pour mesurer les variations spatiales du coût des services du logement selon la ville. Pour tenir compte des variations de qualité selon la région, on exécute des régressions hédoniques semi-logarithmiques distinctes pour construire les indices interrégionaux de Laspeyres, Paasche et Fisher-Törnqvist.

Table des matières

1. Introduction **1**

2. Méthodologie **2**
Équivalence locative 2
Approche hédonique 3

3. Sources des données **4**

4. Modèle statistique **4**
Variables 6
Estimation et indices 7

5. Analyse des données et résultats **8**

6. Conclusion **13**

Bibliographie **14**

Tableaux

Tableau 1 – Estimation du modèle 10
Tableau 2 – Indices de prix du logement spatiaux selon la région métropolitaine de recensement 12

Annexes

Annexe 1 – Sources des données 16
Annexe 2 – Description de l'échantillon 20
Annexe 3 – Spécification et estimation du modèle 26
Annexe 4 – Questionnaire sur le loyer de l'EPA 34

1. Introduction

La Division des prix produit les indices comparatifs des prix de détail entre villes (ou indices de prix spatiaux) depuis 1955. S'ils ont certaines caractéristiques en commun avec l'Indice des prix à la consommation (IPC), mieux connu, ces indices n'en demeurent pas moins assez différents en ce qui concerne leur utilisation et leur construction. De surcroît, leur niveau de détail reste inférieur à celui de l'IPC. Ainsi, en 1995, en vue d'améliorer la couverture des indices spatiaux, la Division a ajouté au programme des catégories de produits, comme les repas au restaurant, l'habillement et le mobilier. Cependant, les coûts liés au logement restaient exclus. Pourtant, il ne faut pas sous-estimer l'importance de cette composante manquante. Les économistes savent depuis longtemps que le coût du logement est le déterminant principal des différences interrégionales de niveau des prix.

La demande de mesures offrant une comparaison géographique entre les niveaux de prix ne cesse de croître. Les différences de prix interrégionales ont une incidence sur les comparaisons interrégionales du coût de la vie. Les données sur ces différences interrégionales de prix sont importantes non seulement pour les administrations publiques et le milieu universitaire, mais aussi pour les entreprises et les particuliers qui envisagent un déménagement. Puisqu'elle quantifie ces écarts, la nouvelle source d'information susmentionnée facilitera éventuellement les prises de décision, car on pourrait l'utiliser comme déflateur pour rajuster les salaires ou les niveaux de revenu régionaux, une question importante dans le contexte socioéconomique actuel.

L'expansion du programme courant des indices de prix spatiaux afin d'y inclure le logement fait depuis longtemps l'objet de recherche à la Division des prix. L'inclusion du logement dans l'indice spatial est essentielle, car la plupart de la variation du coût de la vie au Canada est due à des différences de coût du logement[1]. Cependant, jusqu'à présent la mesure du coût du logement posait des difficultés au programme des indices spatiaux. Plus précisément, l'hétérogénéité du logement empêchait de faire des comparaisons interrégionales sensées des prix[2]. Dans le présent document, nous poursuivons la recherche en poussant plus loin l'élaboration des techniques employées antérieurement et en concentrant notre attention sur un modèle particulier utilisé pour calculer des indices de prix du logement spatiaux expérimentaux[3].

La présente étude s'appuie sur une méthode d'équivalence locative pour produire une mesure du coût des services de logement pour les propriétaires. Afin de tenir compte de la nature hétérogène du logement, nous produisons les indices de Laspeyres, de Paasche et de Fisher-Törnqvist pour la comparaison spatiale des coûts du logement selon une méthode hédonique. Nous utilisons une spécification semi-logarithmique du modèle pour exécuter la régression hédonique corrigée pour la qualité, qui fournit l'information requise pour le calcul des

[1] De plus, il est intéressant de souligner que le logement est la composante principale des dépenses familiales, représentant 27,9 % des dépenses moyennes des ménages en 1992. Étant donné les écarts importants entre les niveaux de prix des logements, la variation du prix des logements peut avoir un effet important sur la répartition budgétaire du consommateur.

[2] Par hétérogénéité du logement, on entend la variété ou diversité du bien « logement ». Le logement englobe une gamme de caractéristiques qualitatives dont nous discutons de façon plus détaillée dans le corps du rapport. Les unités de logement ne contiennent pas toutes les mêmes caractéristiques et, par conséquent, les logements ne sont pas tous égaux ni homogènes.

[3] Les techniques antérieures sont décrites dans Prud'homme et Thivierge [1999].

indices interrégionaux des prix des logements locatifs. Nous calculons 16 indices spatiaux pour 16 centres urbains afin de pouvoir comparer les prix de location corrigés pour la qualité dans les régions métropolitaines de recensement (RMR) du Canada.

À la section 2 du document, nous donnons un aperçu de la méthodologie et des hypothèses qui sous-tendent l'analyse. À la section 3, nous décrivons brièvement les sources de données utilisées. À la section 4, nous présentons le modèle et les méthodes d'estimation, ainsi que le raisonnement qui a mené à leur choix. À la section 5, nous présentons l'analyse des données et les résultats de la régression ainsi que les indices. Enfin, à la section 6, nous présentons certaines conclusions.

2. Méthodologie

Équivalence locative

Le coût des services de logement englobe à la fois les logements occupés en location et ceux occupés en propriété. La méthode généralement reconnue d'évaluation des coûts pour les logements occupés en location est une mesure fondée sur le loyer. Par contre, pour les logements occupés en propriété, la question est controversée et il existe plusieurs méthodes de mesure du coût du logement fondées sur les concepts des décaissements, du prix d'achat net, de l'équivalence locative ou du coût d'utilisation.

La méthode fondée sur les décaissements consiste à calculer le coût du logement comme étant la somme des décaissements réels, y compris ou non compris les paiements au titre de l'avoir propre foncier que doit faire le propriétaire pour utiliser son logement. La méthode fondée sur le prix d'achat net consiste à calculer ce que coûte directement au propriétaire l'achat du logement, en tenant compte des frais d'entretien de ce logement.

Dans la présente étude, nous considérons les méthodes fondées sur l'équivalence locative et sur le coût d'utilisation comme étant les plus pertinentes. La première consiste à déterminer le montant auquel le propriétaire serait prêt à louer le logement et la seconde est basée sur le coût de renonciation lié au fait d'être propriétaire du logement et de l'occuper[4]. Ces méthodes sont opérationnellement différentes, mais conceptuellement équivalentes.

La méthode de l'équivalence locative est celle que nous préférons pour plusieurs raisons. Premièrement, les estimations du coût d'utilisation et le loyer du marché n'évoluent pas parallèlement. En général, les loyers sont assez stables au cours du temps, tandis que les estimations de coût ont tendance à varier. Deuxièmement, la définition et la mesure de variables pertinentes que nécessite l'application de la méthode du coût d'utilisation sont problématiques[5]. Enfin, certains ont soutenu que la collecte de données régionales sur les prix des logements est

[4] Les deux méthodes donnent une mesure du coût du flux de services résultant de la propriété et de l'occupation d'un logement.

[5] Darrough [1983] soutient que les méthodes fondées sur le coût d'utilisation ne devraient pas être jugées non fiables si les variables sont définies et mesurées correctement. Habituellement, les estimations du coût d'utilisation sont sensibles aux diverses hypothèses qui sous-tendent le calcul de la valeur de la variable. Donc, il est essentiel d'estimer correctement les variables pertinentes qui visent à tenir compte des gains en capital et des régimes fiscaux différents appliqués aux biens et aux particuliers.

coûteuse, si bien que la méthode fondée sur le coût d'utilisation exige plus de ressources que la tâche plus simple consistant à recueillir des données sur les loyers du marché. Si la méthode d'équivalence locative n'est pas nécessairement la meilleure, elle est certainement la plus pratique pour le calcul des estimations spatiales [Darrough, 1983].

La méthode de l'équivalence locative consiste à utiliser les loyers observés sur le marché comme approximation du coût de propriété et, de façon plus générale, du logement. Elle est fondée sur la question de savoir combien le propriétaire demanderait pour louer sa maison à une personne telle que lui [Darrough, 1983; 603]. L'observation des prix de location réels du marché permet d'estimer le loyer éventuel. Cette méthode s'appuie sur l'hypothèse que le coût du logement ou du service de logement subit par les propriétaires peut être représenté par les loyers observés sur le marché. Par conséquent, la méthode d'équivalence locative est celle que nous préférons ici et que nous employons dans l'analyse qui suit.

Approche hédonique

L'approche de la régression hédonique est utilisée de longue date en économie empirique pour évaluer les rapports qualité-prix. Elle est fondée essentiellement sur le concept utilitariste consistant à comparer les prix des biens et services en vue d'évaluer le bien-être et la satisfaction que le consommateur tire de leur consommation. Dans le contexte de l'approche hédonique, on considère la plupart des biens et services comme étant un produit composite formé d'une série d'autres biens et services habituellement appelés attributs ou caractéristiques qualitatives. La contribution de chaque attribut à la valeur du produit composite peut être calculée par des méthodes hédoniques. Donc, on peut appliquer la méthode de régression hédonique aux logements en vue de mesurer les effets de certaines de leurs caractéristiques qualitatives sur le prix de location. Par conséquent, si l'on considère deux logements locatifs échangés sur le même marché et que le premier a des caractéristiques plus désirables que le second, le premier sera jugé de plus grande valeur par les locataires et aura un prix de location plus élevé[6].

La méthode hédonique reconnaît le prix du logement comme étant le prix d'une série de biens et de services qui sont regroupés pour donner un produit composite appelé « logement ». La série de biens et de services sont les caractéristiques qualitatives qui constituent l'espace pur. Bien que les caractéristiques qualitatives du logement ne soient pas vendues séparément, l'utilisation de la méthode hédonique permet de calculer un prix implicite pour chaque attribut qui forme le produit composite. Il en est ainsi parce qu'on émet l'hypothèse qu'il existe des marchés implicites pour ces caractéristiques qualitatives. On suppose que les prix implicites sont déterminés par l'offre ou générés par le marché et extérieurs aux décisions de consommation du ménage (comportement d'acceptation du prix du marché). L'hypothèse selon laquelle les prix sont déterminés par l'offre est valide si le bien en question est un bien durable (comme un logement) dont l'offre est assez stable [Muellbauer, 1974]. Cependant, l'approche utilisée dans la présente étude est axée sur la demande, car nous considérons le bien-être des consommateurs et la méthode hédonique ne doit donc pas viser à expliquer complètement les phénomènes du marché.

[6] On peut interpréter l'expression « plus désirable » comme étant non pas la qualité de l'attribut, mais la quantité de la caractéristique qualitative.

Il découle des hypothèses susmentionnées qu'on peut décomposer le prix de location en prix marginaux des attributs. Cependant, pour y arriver, il faut aussi supposer que les caractéristiques qualitatives sont homogènes et que les biens formant les caractéristiques qualitatives composites sont, eux aussi, homogènes. Cette hypothèse implique que l'état courant est le court terme et que tout changement de caractéristique et de qualité ne peut avoir lieu que dans le long terme. L'hypothèse d'homogénéité est également importante parce qu'elle permet aux biens d'avoir une nature objective. Par conséquent, la décision de louer prise par le consommateur repose uniquement sur la série de caractéristiques et non sur la qualité des attributs particuliers. Il est donc possible d'isoler l'effet des changements de qualité sur les prix [Lancaster, 1966].

Par exemple, dans le cas des chambres à coucher, l'hypothèse que les caractéristiques sont homogènes implique que toutes les chambres sont identiques. Aucune différence n'existe en ce qui a trait au nombre de pieds carrés, au nombre de fenêtres, à la taille des placards ou à l'existence éventuelle d'une salle de bain attenante. Donc, l'hypothèse émise est que le consommateur décide de louer un logement particulier en s'appuyant strictement sur la « série » de caractéristiques qualitatives et non sur la « qualité » des caractéristiques qualitatives. Pour en revenir à l'exemple des chambres à coucher, le consommateur prend sa décision strictement d'après le nombre de chambres à coucher dans le logement et non d'après les caractéristiques qualitatives de chaque chambre à coucher (c.-à-d. la taille de la chambre à coucher, le nombre de fenêtres, la taille du placard ou l'accès à une salle de bain). Brièvement, on suppose que toutes les chambres à coucher sont identiques dans un logement, ainsi que d'un logement à l'autre et d'un emplacement à l'autre. Autrement dit, une chambre à coucher est la même à Montréal qu'à Regina, et il en est ainsi pour tous les attributs qui constituent le bien « logement ».

3. Sources des données

Les deux sources principales de données utilisées pour l'analyse sont l'Enquête sur la population active (EPA) [août 1998 à décembre 1999] et le Recensement du Canada de 1996. La taille de l'échantillon est de 14 434 ménages. Le questionnaire sur le loyer de l'EPA fournit les caractéristiques physiques qualitatives des logements et le Recensement fournit les caractéristiques socioéconomiques des quartiers dans lesquels sont situés les logements. Par conséquent, l'unité primaire d'analyse est le logement et l'unité secondaire, le quartier ou le secteur de dénombrement (SD). Une description plus détaillée de la population cible, du plan de sondage et de la façon dont les sources de données sont intégrées figure à l'annexe 1. Pour la taille de l'échantillon selon la région urbaine, consulter l'annexe 2.

4. Modèle statistique

La théorie économique ne propose aucune forme fonctionnelle pour les modèles hédoniques. Habituellement, les chercheurs s'appuient sur la littérature et les critères de qualité de l'ajustement du modèle pour choisir une forme fonctionnelle appropriée. Prud'homme et Thivierge [1999] ont passé en revue plusieurs modèles qui pourraient être utilisés pour la construction d'indices hédoniques. Plus précisément, des modèles de régressions distinctes, des modèles à variables fictives et des modèles à coefficients aléatoires ont été appliqués sous diverses spécifications, dont les formes linéaire, semi-logarithmique et log-linéaire.

Pour défendre le choix de la méthode des régressions distinctes, Muellbauer [1974] soutient que, dans la mesure du possible, les marchés devraient être segmentés en se fondant sur des groupes de produits tels qu'il soit probable que les consommateurs aient un taux marginal de substitution (TMS) comparable et que les segments obtenus devraient être étudiés séparément. Il ajoute que cette segmentation réduit les variations transversales, mais au prix d'un accroissement des problèmes de multicollinéarité. Dans le contexte spatial du logement, l'approche des régressions distinctes implique qu'on exécute une régression pour chaque segment, région métropolitaine de recensement ou emplacement, au lieu de regrouper tous les segments ou emplacements, de tenir compte de l'effet de la région ou du segment et d'exécuter une régression unique. Dans le présent document, nous supposons que le TMS est le même pour tous les consommateurs dans un emplacement ou un segment donné et, par conséquent, nous étudions séparément ces régions.

Comme l'ont fait Prud'homme et Thivierge [1999] et d'autres chercheurs, dans la présente étude, nous limitons les modèles hédoniques du loyer à celui des régressions distinctes, puis nous employons deux tests de spécification pour repérer celle qui produira le meilleur ajustement compte tenu des données disponibles (annexe 3).

Les études hédoniques réalisées par MacDonald [1986] et Salois [1987] s'appuient aussi sur des spécifications linéaires et semi-logarithmiques pour étudier le prix des logements. Dans leur étude en simulation, Cropper, Deck et McConnell [1988] comparent la constance des propriétés d'une large gamme de modèles auxquels sont appliqués divers scénarios de mesure et d'omission de variables et constatent que les formes fonctionnelles plus simples, comme les modèles linéaire, semi-logarithmique ou bilogarithmique et le modèle linéaire de Box-Cox donnent de nettement meilleurs résultats que le modèle quadratique ou le modèle quadratique de Box-Cox.

Le choix d'une spécification devrait se fonder sur les propriétés inhérentes de cette dernière. Mais, avant tout, la spécification devrait avoir une utilité analytique *a priori* et être un descripteur exact du caractère sous-jacent des données. Toute spécification qui donne à penser que la relation entre les variables dépendante et indépendantes pourrait être représentée par une spécification plus appropriée devrait être rejetée.

Nous utilisons ici un modèle semi-logarithmique pour décrire la relation entre les prix de location et les attributs des logements. Il est évident, d'après les tests, que la spécification semi-logarithmique est le meilleur descripteur des données. L'utilité analytique *a priori* du modèle tient à trois éléments importants : le modèle a donné de bons résultats lors des tests d'évaluation de la qualité de l'ajustement décrit dans la littérature [Moulton, 1995], il a été démontré qu'il est facile de l'utiliser pour la production routinière d'indices [Prud'homme et Yu, 1999] et ses coefficients de régression peuvent être interprétés. Les coefficients explicatifs (β) mesurent la variation proportionnelle, ou relative, constante de la variable dépendante (Y) pour une variation absolue donnée de la variable indépendante (X).

$$\beta = \frac{\Delta \textit{ relative de } Y}{\Delta \textit{ absolue de } X} \cong \frac{\dfrac{Y_k - Y_{k-1}}{Y_{k-1}}}{X_k - X_{k-1}} \qquad (1)$$

Donc, les coefficients exogènes varient proportionnellement à la variation de la variable endogène.

Nous procédons à la régression du logarithme naturel du prix de location sur les attributs physiques et socioéconomiques selon l'emplacement. Le modèle semi-logarithmique des régressions distinctes peut s'écrire sous la forme :

$$Y_{vk} = \alpha_v + \sum_{j=1}^{J} \beta_{vj} X_{vjk} + \varepsilon_{vk} \qquad (2)$$

où Y_{vk} est le logarithme naturel du loyer mensuel pour le k^e logement dans l'emplacement v; α_v est la coordonnée à l'origine pour l'emplacement v; β_{vj} est le vecteur de coefficients pour les j caractéristiques qualitatives des logements et X_{vjk} est la matrice de tous les attributs (physiques et socioéconomiques). Le terme d'erreur est représenté par ε_{vk} pour le k^e logement dans l'emplacement v.

Variables

Essentiellement, il existe trois modes de répartition des données : physique, socioéconomique et spatiale. Les régressions distinctes selon l'emplacement tiennent compte des divisions spatiales. Les attributs des logements, ou variables exogènes, comprennent les caractéristiques physiques qualitatives et les attributs socioéconomiques.

Les caractéristiques physiques qualitatives sont des attributs propres à un logement individuel. Un logement est « un ensemble de pièces d'habitation distinct au niveau de la structure des pièces d'habitation d'autres logements, et possédant une entrée privée avec accès sur l'extérieur ou sur un corridor ou un escalier commun à l'intérieur » [Guide de l'Enquête sur la population active, 2000]. Les caractéristiques physiques définissent la nature du logement, donc le flux de services que fournit ce logement. Les attributs socioéconomiques sont particuliers au secteur de dénombrement (SD) ou au secteur de recensement (SR). Par conséquent, les logements compris dans un même secteur de dénombrement ont des attributs socioéconomiques identiques. Les caractéristiques socioéconomiques sont exprimées en pourcentage de logements dans le SD possédant l'attribut en question.

Les caractéristiques physiques incluses dans la régression à titre de variables indépendantes sont le type de logement, l'âge de la structure, le nombre de chambres à coucher (du studio sans chambre à coucher au logement à quatre chambres à coucher ou plus) et une série de produits sous forme de variables fictives. Ces produits représentent les biens et services compris dans le prix de location, c'est-à-dire le chauffage, l'eau chaude, l'électricité, le réfrigérateur, la cuisinière, la laveuse, la sécheuse, la câblodistribution, le mobilier, d'autres gros appareils ménagers et le stationnement. Certaines variables composites de produits sont créées pour limiter l'effet de la multicollinéarité. Le calcul des variables composites est décrit plus en détail à l'annexe 3 et toutes les variables utilisées dans les régressions distinctes sont énumérées à l'annexe 1.

Les variables socioéconomiques sont incluses dans le modèle pour éviter le problème de l'approximation de variable. L'intégration de renseignements sur le quartier a pour but de construire un modèle dont la spécification est plus complète. Les variables indépendantes de quartier utilisées sont le pourcentage de logements privés loués dans un secteur de dénombrement, le pourcentage de logements privés qui ont besoin de grosses réparations, le pourcentage de personnes ayant fait des études postsecondaires, la distance moyenne à parcourir pour se rendre au travail (km) et le pourcentage de personnes vivant sous le seuil de faible revenu. Les seuils de faible revenu sont établis pour des niveaux de revenu différents selon la taille de la région de résidence (ou le degré d'urbanisation) et la taille de la famille. On utilise les données nationales de l'Enquête sur les dépenses des familles de 1992 pour calculer le seuil de faible revenu et l'Indice des prix à la consommation pour mettre les niveaux à jour annuellement. Nous considérons que les familles consacrant au moins 70 % de leurs revenus aux nécessités de la vie, comme les aliments, le logement et les vêtements vivent dans des circonstances « difficiles » et les classons dans la catégorie des familles à faible revenu. Le seuil est un nombre arbitraire utilisé pour établir la démarcation entre le groupe d'unités familiales « à faible revenu » et le groupe des « autres familles ». Toute famille dont le revenu est égal ou supérieur au seuil est considérée comme appartenant à la catégorie « autre » [Dictionnaire du recensement de 1996 (1997)]. En incluant les caractéristiques socioéconomiques, nous supposons qu'il existe des marchés implicites pour ces caractéristiques qualitatives ambiantes régionales et de quartier.

Estimation et indices

Le calcul des indices hédoniques se fait en deux étapes. On commence par estimer les coefficients des caractéristiques qualitatives au moyen d'une régression du prix de location sur les caractéristiques qualitatives. Puis, on se sert des coefficients obtenus pour construire un indice de prix spatial de Fisher-Törnqvist. Avant d'exécuter les régressions, il faut résoudre les problèmes que posent la spécification du modèle, la multicollinéarité, les variables omises et les erreurs non sphériques. Ces questions sont décrites en détail à l'annexe 3.

La première étape de l'estimation des indices de prix rajustés pour la qualité consiste à générer des coefficients de caractéristiques qualitatives en exécutant une régression distincte pour chaque région pour laquelle un indice est calculé (c.-à-d. chaque centre urbain). En outre, nous utilisons un estimateur par les moindres carrés généralisés réalisables (MCGR) en deux étapes pour tenir compte de la variation des erreurs aléatoires à l'intérieur des classes d'attributs[7].

La deuxième étape de la production des indices interrégionaux consiste à appliquer les coefficients susmentionnés à la construction d'un indice de prix spatial de Fisher-Törnqvist (*FT*) qui est une moyenne géométrique des indices de prix de Laspeyres (*L*) et de Paasche (*P*). Gillingham [1975] a consolidé la pratique consistant à construire une moyenne géométrique au moyen d'équations hédoniques semi-logarithmiques rajustées pour la qualité et Moulton [1995] a étendu l'utilisation des équations hédoniques aux indices interrégionaux. L'indice de Laspeyres s'appuie sur la valeur moyenne de la caractéristique j de la région de référence comme base (X_{jb}) pour pondérer la différence entre les coefficients régionaux :

[7] Une discussion de l'estimateur utilisée dans la présente analyse figure à l'annexe 3.

$$P^{L}_{v_i v_o} = e^{\sum_j \left(\hat{\beta}_{jv_i} - \hat{\beta}_{jv_o} \right) \left(X_{jv_o} \right)} \qquad (3)$$

Mentionnons que la région de référence est représentée par *o* et la région comparée par *i*. L'indice de Paasche s'appuie sur la valeur moyenne de la caractéristique *j* de la région comparée comme base ($\bar{X}_{jv_i}$) :

$$P^{P}_{v_i v_o} = e^{\sum_j \left(\hat{\beta}_{jv_i} - \hat{\beta}_{jv_o} \right) \left(X_{jv_i} \right)} \qquad (4)$$

Si nous calculons la moyenne géométrique des deux indices, nous obtenons un indice de Fisher-Törnqvist :

$$P^{FT}_{v_i v_o} = \sqrt{P^{L}_{v_i v_o} P^{P}_{v_i v_o}} = \sqrt{e^{\sum_j \left(\hat{\beta}_{jv_i} - \hat{\beta}_{jv_o} \right) \left(X_{jv_o} \right)} e^{\sum_j \left(\hat{\beta}_{jv_i} - \hat{\beta}_{jv_o} \right) \left(X_{jv_i} \right)}} \qquad (5)$$

$$P^{FT}_{v_i v_o} = \left(P^{L}_{v_i v_o} P^{P}_{v_i v_o} \right)^{\frac{1}{2}} = e^{\frac{1}{2}\left[\sum_j \left(\hat{\beta}_{jv_i} - \hat{\beta}_{jv_o} \right) \left(X_{jv_o} + X_{jv_i} \right) \right]} \qquad (6)$$

L'utilisation d'une moyenne géométrique des indices de Laspeyres et de Paasche signifie que les valeurs moyennes de la caractéristique *j* de ces indices sont utilisées pour la région de référence et la région comparée, respectivement. Le problème de la non-transitivité de l'indice de Fisher-Törnqvist persiste[8].

5. Analyse des données et résultats

L'ajustement global du modèle est généralement bon et conforme à celui observé lors d'études transversales antérieures. Si la majorité des coefficients de variables ont le signe correct *a priori* et sont d'un ordre de grandeur raisonnable, on suppose alors que tout écart par rapport aux informations *a priori* est causé par les variables omises. Les variables incluses dans le modèle pourraient servir d'approximation si le modèle est spécifié incorrectement ou ne comprend pas suffisamment d'attributs du bien en question pour invalider les relations d'approximation.

Nous nous attendions à ce que les variables discrètes d'âge aient un coefficient négatif. En général, la volonté de payer pour un espace diminue à mesure que l'âge de la structure augmente. Nous nous attendions aussi à ce que la variable de type de logement (DYTP1234) ait un signe positif puisqu'en principe les locataires préfèrent vivre dans une maison unifamiliale qu'en appartement. L'argument à l'appui est que, s'ils ont l'impression que les taux de mortalité, de criminalité, de maladie mentale, de délinquance juvénile et de surpeuplement des logements sont plus faibles dans les quartiers à maisons unifamiliales, les locataires pourraient être disposés

[8] Les propriétés de ces trois indices sont décrites de façon plus détaillée à l'annexe 3.

à payer une prime pour ce genre de logement et le coefficient de DYTP1234 devrait être positif[9]. Nous nous attendions à ce que les chambres à coucher et toutes les variables de produits aient un coefficient positif, car elles ajoutent du confort au logement. Le signe de la majorité des coefficients correspond à nos attentes. Ceux pour lesquels il n'en est pas ainsi sont généralement non significatifs (voir le tableau 1). Donc, nous considérons que la question des variables omises ne pose pas de problèmes graves en ce qui concerne la spécification du modèle.

Reste à examiner la question du signe improbable des estimations (signe positif ou négatif improbable) et des erreurs-types importantes pour certains coefficients. Elle pourrait tenir à la fréquence (forte ou faible) d'une variable donnée dans la ville considérée. Quand la fréquence d'une variable est très faible, comme dans le cas des laveuses et des sécheuses à Montréal, ou très élevée, comme dans le cas des cuisinières et des réfrigérateurs à Charlottetown, il devient impossible d'estimer exactement le prix implicite des biens. Comme l'a suggéré Moulton [1995], le rajustement de certains marchés ne se fait pas par la voie du prix implicite d'une caractéristique qualitative, mais par celle des quantités consommées. Ce résultat appuie l'utilisation de moyennes régionales pour les caractéristiques qualitatives dans le calcul des indices. Si l'on utilise des moyennes nationales, les conditions limites, comme les exemples susmentionnés, pourraient être caractérisées incorrectement.

Le pourcentage de logements privés loués dans un secteur de dénombrement et le pourcentage de personnes ayant fait des études postsecondaires devraient, en principe, avoir un signe positif. Nous avons supposé que les locataires choisissent sélectivement de s'associer à des personnes partageant les mêmes valeurs et motivations qu'eux. Donc, ils seront disposés à payer une prime pour être proches de locataires ayant les mêmes niveaux d'études et de revenus. Le pourcentage de logements privés ayant besoin de grosses réparations, la distance moyenne à parcourir pour se rendre au travail et le seuil de faible revenu sont des variables qui, théoriquement, devraient toutes avoir un signe négatif. Nous nous attendions à ce que les locataires ne préfèrent pas vivre dans un logement en état de délabrement, donc que la location se fasse au rabais dans les quartiers où la proportion de logements nécessitant des réparations importantes est élevée. Nous nous attendions aussi à ce que le loyer payé pour un logement diminue à mesure que la distance à parcourir pour se rendre au travail augmente; donc, les économies dues à une plus courte distance à parcourir pour se rendre au travail devraient être intégrées sous forme de prime dans le prix total de location. En principe, la variable de seuil de faible revenu devrait avoir un coefficient négatif, car il devrait exister une corrélation positive entre le niveau de revenu et le montant du loyer payé. L'effet des caractéristiques du quartier est habituellement significatif; cependant, il ne concorde pas systématiquement avec nos attentes *a priori* concernant le signe (voir le tableau 1).

[9] Cependant, il existe des théories concurrentes. Le signe du coefficient de la variable de type de logement doit être négatif si l'on pense que le niveau de densité urbaine influence le choix du locataire. Ainsi, si les locataires visent davantage à réduire leur rôle dans l'entretien du logement (jardinage, déneigement, etc.), alors ils pourraient accorder à la possibilité d'abandonner ces activités une valeur implicite qui seraient intégrées sous forme de prime dans le prix de location d'un appartement. Certains avantages sociaux liés à la vie dans une zone fortement peuplée, comme des occasions sans égal de gratification, un surplus de possibilités et de stimulations, et l'exposition non délibérée à l'éducation, au cosmopolitisme et aux idées novatrices, pourraient aussi faire accepter une prime pour vivre en appartement [Choldin, 1978].

Tableau 1

Estimation du modèle

Variable dépendante : Log (loyer payé)

Variable	St. John's		Charlottetown		Halifax		Saint John	
	Coefficients	Err.-type	Coefficients	Err.-type	Coefficients	Err.-type	Coefficients	Err.-type
Coordonnée à l'origine	5,609	0,146***	5,577	0,079***	5,640	0,071***	5,432	0,142***
10 < âge < 20 ans	0,001	0,053	-0,094	0,017***	0,087	0,023***	-0,223	0,066***
20 < âge < 40 ans	-0,024	0,051	-0,121	0,016***	0,039	0,015***	-0,211	0,061***
> 40 ans	0,113	0,066*	-0,112	0,018***	0,028	0,025	-0,159	0,065***
Maison unifamiliale	-0,088	0,046*	0,022	0,017	-0,063	0,027***	0,002	0,055
Chambre à coucher	0,100	0,021***	0,194	0,008***	0,194	0,008***	0,151	0,014***
Produits de luxe	0,135	0,132	0,026	0,015*	0,187	0,016***	-0,035	0,121
Services publics	0,053	0,112	0,063	0,030**	0,054	0,017***	0,027	0,029
Électricité	0,010	0,114	0,064	0,017***	0,003	0,027	0,049	0,030
Réfrigérateur/ cuisinière	0,022	0,065	0,049	0,057	0,088	0,049*	0,134	0,086
Laveuse/sécheuse	0,042	0,072	0,096	0,033***	0,064	0,040	0,065	0,074
Stationnement	0,055	0,055	0,040	0,027	-0,096	0,016***	0,171	0,032***
% de logements loués	0,002	0,001***	0,001	0,000***	0,001	0,000***	0,005	0,000***
% avec grosses réparations	0,001	0,004	-0,002	0,001	0,003	0,002**	-0,008	0,003***
% études postsecondaires	0,004	0,001***	0,003	0,001***	0,005	0,001***	0,001	0,001
Distance à parcourir	0,007	0,007	-0,009	0,003***	-0,007	0,001***	0,003	0,004
% sous le SFR	-0,003	0,001*	0,001	0,001	-0,004	0,001***	-0,002	0,001***
R^2 corr.	**0,255**		**0,643**		**0,892**		**0,714**	

Variable	Fredericton		Québec		Montréal		Hull	
	Coefficients	Err.-type	Coefficients	Err.-type	Coefficients	Err.-type	Coefficients	Err.-type
Coordonnée à l'origine	5,679	0,147***	5,928	0,059***	5,732	0,040***	5,761	0,052***
10 < âge < 20 ans	-0,028	0,064	-0,076	0,041*	-0,122	0,033***	-0,128	0,022***
20 < âge < 40 ans	-0,084	0,060	-0,223	0,030***	-0,147	0,028***	-0,177	0,023***
> 40 ans	-0,177	0,067***	-0,314	0,036***	-0,228	0,028***	-0,137	0,026***
Maison unifamiliale	-0,066	0,046	-0,038	0,020*	0,023	0,018	0,029	0,019
Chambre à coucher	0,174	0,018***	0,180	0,011***	0,135	0,005***	0,182	0,007***
Produits de luxe	0,168	0,046***	0,064	0,057	0,085	0,037**	0,028	0,023
Services publics	0,105	0,054*	0,115	0,021***	0,160	0,009***	0,025	0,026
Électricité	0,087	0,067	-0,018	0,021	-0,011	0,019	0,089	0,027***
Réfrigérateur/ cuisinière	-0,233	0,071***	-0,112	0,024***	-0,088	0,011***	0,065	0,014***
Laveuse/sécheuse	0,151	0,055***	0,031	0,049	0,106	0,089	0,065	0,048
Stationnement	0,096	0,063	-0,055	0,030*	-0,001	0,009	0,030	0,027
% de logements loués	0,006	0,001***	0,003	0,000***	0,003	0,000***	0,001	0,000***
% avec grosses réparations	0,010	0,003***	-0,009	0,002***	-0,008	0,001***	-0,002	0,002
% études postsecondaires	0,006	0,001***	0,002	0,001***	0,005	0,000***	0,003	0,001***
Distance à parcourir	-0,025	0,007***	-0,004	0,003	0,000	0,002	-0,010	0,004**
% sous le SFR	-0,010	0,002***	-0,004	0,001***	-0,003	0,000***	-0,002	0,001***
R^2 corr.	**0,598**		**0,470**		**0,372**		**0,730**	

*** Valeur statistiquement significative au niveau de signification de 1 %.

** Valeur statistiquement significative au niveau de signification de 5 %.

* Valeur statistiquement significative au niveau de signification de 10 %.

Tableau 1 (suite)
Estimation du modèle

Variable dépendante : Log (loyer payé)

Variable	Ottawa		Toronto		Winnipeg		Regina	
	Coefficients	Err.-type	Coefficients	Err.-type	Coefficients	Err.-type	Coefficients	Err.-type
Coordonnée à l'origine	6,031	0,060***	6,110	0,049***	6,211	0,077***	6,160	0,145***
10 < âge < 20 ans	-0,200	0,039***	-0,015	0,027	0,037	0,066	-0,518	0,130***
20 < âge < 40 ans	-0,224	0,033***	-0,086	0,027***	-0,086	0,066	-0,598	0,129***
> 40 ans	-0,260	0,034***	-0,143	0,028***	-0,142	0,067**	-0,672	0,129***
Maison unifamiliale	0,026	0,021	-0,020	0,018	-0,074	0,022***	-0,080	0,024***
Chambre à coucher	0,183	0,008***	0,173	0,005***	0,112	0,007***	0,151	0,011***
Produits de luxe	0,107	0,023***	0,008	0,016	0,094	0,011***	0,008	0,016
Services publics	0,043	0,016***	-0,003	0,016	0,032	0,012***	0,034	0,021
Électricité	0,042	0,017**	0,041	0,012***	-0,021	0,013*	-0,017	0,038
Réfrigérateur/ cuisinière	0,049	0,032	-0,037	0,031	-0,091	0,035***	0,056	0,056
Laveuse/sécheuse	0,009	0,033	0,136	0,018***	0,090	0,017***	0,059	0,020***
Stationnement	0,060	0,014***	0,074	0,010***	0,061	0,013***	0,070	0,033***
% de logements loués	0,001	0,000***	0,001	0,000***	0,000	0,000**	0,000	0,000
% avec grosses réparations	-0,001	0,001	-0,005	0,001***	-0,007	0,001***	-0,006	0,002***
% études postsecondaires	0,005	0,000***	0,005	0,000***	0,001	0,000***	0,004	0,001***
Distance à parcourir	-0,007	0,002***	-0,004	0,002**	-0,018	0,002***	-0,001	0,003
% sous le SFR	-0,003	0,000***	0,001	0,000***	-0,003	0,000***	0,001	0,001*
R^2 corr.	**0,631**		**0,477**		**0,513**		**0,650**	

Variable	Calgary		Edmonton		Vancouver		Victoria	
	Coefficients	Err.-type	Coefficients	Err.-type	Coefficients	Err.-type	Coefficients	Err.-type
Coordonnée à l'origine	5,788	0,195***	6,199	0,145***	5,644	0,055***	5,694	0,129***
10 < âge < 20 ans	0,029	0,168	-0,296	0,128**	-0,174	0,030***	-0,073	0,103
20 < âge < 40 ans	-0,133	0,168	-0,373	0,127***	-0,103	0,027***	-0,237	0,087***
> 40 ans	-0,137	0,170	-0,483	0,130***	-0,207	0,028***	-0,090	0,088
Maison unifamiliale	0,079	0,028***	0,103	0,023***	-0,009	0,022	0,063	0,033*
Chambre à coucher	0,186	0,010***	0,142	0,009***	0,221	0,008***	0,208	0,014***
Produits de luxe	0,102	0,019***	0,062	0,019***	-0,073	0,010***	0,271	0,023***
Services publics	0,059	0,027**	0,059	0,021***	0,013	0,016	0,116	0,022***
Électricité	0,002	0,030	0,065	0,024***	-0,132	0,023***	-0,169	0,066***
Réfrigérateur/ cuisinière	0,033	0,074	-0,036	0,060	0,151	0,017***	0,070	0,063
Laveuse/sécheuse	0,078	0,022***	0,028	0,016	0,115	0,021***	-0,018	0,026
Stationnement	0,057	0,027**	-0,028	0,026	0,064	0,013***	-0,041	0,021**
% de logements loués	0,003	0,000***	0,001	0,000***	0,000	0,000	0,003	0,001***
% avec grosses réparations	-0,002	0,002	0,002	0,001	0,000	0,001	-0,005	0,002***
% études postsecondaires	0,003	0,001***	0,001	0,000	0,012	0,001***	0,007	0,001***
Distance à parcourir	-0,017	0,003***	0,005	0,002***	-0,021	0,002***	0,001	0,003
% sous le SFR	-0,002	0,001***	-0,003	0,000***	-0,002	0,001***	-0,005	0,001***
R^2 corr.	**0,681**		**0,550**		**0,837**		**0,716**	

*** Valeur statistiquement significative au niveau de signification de 1 %.

** Valeur statistiquement significative au niveau de signification de 5 %.

* Valeur statistiquement significative au niveau de signification de 10 %.

L'utilisation d'une forme fonctionnelle semi-logarithmique donne des coefficients de régression interprétables. Ces coefficients mesure la variation relative du loyer payé pour une variation absolue donnée d'une variable explicative. Par exemple, on peut facilement évaluer l'impact sur le prix de location d'une augmentation du nombre de chambres à coucher dans un logement, disons, à Winnipeg et à Vancouver. Pour toute chambre à coucher supplémentaire, le prix de location augmente de 22,1 % à Vancouver, mais n'augmente que de 11,2 % à Winnipeg. Ces résultats reflètent la disposition à payer pour une chambre à coucher supplémentaire à Vancouver comparativement à Winnipeg et les primes associées aux pièces supplémentaires dans ces deux centres urbains. Les coefficients des autres variables explicatives peuvent être interprétés de façon comparable.

Les indices de Laspeyres, de Paasche et de Fisher sont présentés pour la RMR de Winnipeg comme ville de référence. À l'annexe 3, nous donnons les indices estimés pour 14 paires de villes. L'indice de Fisher est la moyenne géométrique des indices de Laspeyres et de Paasche et sa valeur devraient systématiquement être comprises entre celles de ces deux indices. L'ordre de grandeur des indices de Laspeyres et de Paasche dépend de la région de référence et de la région comparée, ainsi que de la relativité de la caractéristique moyenne de base. La comparaison entre les indices obtenus et nos attentes est bonne (voir le tableau 2). Nous nous attendions à ce que les indices soient plus élevés pour Ottawa, Toronto et Vancouver que pour les autres centres urbains[10]. Si l'on utilise l'indice de Fisher, les quatre régions urbaines où le coût du logement corrigé pour la qualité est le plus élevé sont Toronto, Ottawa, Vancouver et Victoria. Les quatre villes où ce coût corrigé pour la qualité est le plus faible sont, par ordre croissant, Québec, Saint John, Regina et Charlottetown.

Tableau 2
Indices de prix du logement spatiaux selon la région métropolitaine de recensement
(Indice de base Winnipeg=100)

	LASPEYRES	PAASCHE	FISHER
St. John's	106,7	97,8	102,2
Charlottetown	96,4	95,6	96,0
Halifax	111,9	109,5	110,7
Saint John	84,6	92,9	88,6
Fredericton	105,9	115,9	110,8
Québec	85,9	89,9	87,9
Montréal	101,3	100,2	100,8
Hull	103,0	98,2	100,6
Ottawa	131,6	142,8	137,1
Toronto	162,6	174,2	168,3
Winnipeg	**100,0**	**100,0**	**100,0**
Regina	92,2	93,6	92,9
Calgary	118,5	125,1	121,7
Edmonton	99,6	103,9	101,7
Vancouver	124,5	150,2	136,8
Victoria	121,3	135,8	128,4

[10] Les attentes formulées dans le cadre de l'étude sont fondées sur Prud'homme et Thivierge [1999].

6. Conclusion

Depuis plusieurs années, la Division des prix produit une famille d'indices de prix permettant de comparer les « niveaux » de prix entre les villes. Cependant, jusqu'à présent, leur utilisation était limitée parce que, contrairement à l'IPC apparenté, ces indices spatiaux n'incluent pas le coût du logement, c'est-à-dire la composante la plus importante des dépenses familiales et, souvent, la composante la plus variable. Les comparaisons géographiques du coût de la vie posent généralement des problèmes méthodologiques, mais ceux-ci sont particulièrement épineux dans le cas du logement à cause de la nature hétérogène de ce dernier.

Pour surmonter les difficultés que posent les comparaisons spatiales du logement, nous avons créé une nouvelle base de données en appariant le questionnaire sur le loyer de l'Enquête sur la population active et le questionnaire du Recensement de 1996. La base de données contient des renseignements détaillés sur les logements locatifs. Nous combinons l'approche de l'équivalence locative et une méthode basée sur des régressions hédoniques semi-logarithmiques distinctes rajustées pour la qualité en vue de construire des indices de Laspeyres, de Paasche et de Fisher-Törnqvist. En nous concentrant sur un modèle particulier servant au calcul des indices de prix spatiaux expérimentaux, nous examinons en détail un cadre analytique pratique. Cette méthode est techniquement et analytiquement facile à appliquer et les résultats obtenus sont comparables à ceux publiés antérieurement dans la littérature traitant de ce domaine.

La qualité des indices de prix obtenus est tellement fiable qu'on peut maintenant faire des comparaisons du coût du logement entre villes et construire un indice de prix « d'ensemble » spatial. La série d'indices de prix géographiques étendue comprenant le logement donne un indice « d'ensemble » spatial apparenté à l'IPC qui permettra de répondre à la demande croissante de renseignements de ce genre. Les données pourraient être utiles, par exemple, pour rajuster la rémunération des employés mutés et pour déflater les revenus nominaux dans les études des écarts interrégionaux de revenu.

Bibliographie

Allen, R.G.D. 1975. *Index Numbers in Theory and Practice*. Chicago : Aldine Publishing Company.

Amemiya, Takeshi. 1985. *Advanced Econometrics*. Massachusetts : Harvard University Press.

Carroll, Raymond J. et David Ruppert. 1982a. « Robust Estimation in Heteroscedastic Linear Models ». *Annals of Statistics*. 10(2): 429–441.

Carroll, Raymond J. et David Ruppert. 1982b. « A Comparison Between Maximum Likelihood and Generalized Least Squares in a Heteroscedastic Linear Model ». *Journal of the American Statistical Association*. 77(380): 878–882.

Choldin, Harvey M. 1978. « Urban Density and Pathology ». *Annual Review of Sociology*. 4: 91–113.

Cropper, Maureen L., Leland B. Deck, et Kenneth E. McConnell. 1988. « On the Choice of Functional Form for Hedonic Price Functions ». *The Review of Economics and Statistics*. 70(4): 668–675.

Darrough, Masako N. 1983. « The Treatment of Housing in a Cost-of-Living Index: Rental Equivalence and User Cost », dans *La mesure du niveau des prix : Actes du colloque tenu sous l'égide de Statistique Canada*, édit. W.E. Diewert et C. Montmarquette, p. 599–618. Résumé en français sous le titre « Le traitement du logement dans un indice du coût de la vie : Équivalence locative et coût d'utilisation ». Ottawa.

Deaton, Angus et John Muellbauer. 1980. *Economics and Consumer Behaviour*. Cambridge : Cambridge University Press.

Gambino, J.G., M.P. Singh, J. Dufour, D. Kennedy, et J. Lindeyer. 1998. *Méthodologie de l'Enquête sur la population active du Canada*. Statistique Canada. Division des méthodes des enquêtes auprès des ménages. Ottawa : Ministre de l'Industrie. N° 71-526-XPB au catalogue.

Gillingham, Robert F. 1975. « Place to Place Rent Comparisons Using Hedonic Quality Adjustment Techniques ». *Annals of Economic and Social Measurement*. 4: 153–174.

Godfrey, L.G., Michael McAleer, et C.R. McKenzie. 1988. « Variable Addition and Lagrange Multiplier Tests for Linear and Logarithmic Regression Models ». *The Review of Economics and Statistics*. 70(3): 492–503.

Jobson, J.D. et W.A. Fuller. 1980. « Least Squares Estimation When the Covariance Matrix and Parameter Vector are Functionally Related ». *Journal of the American Statistical Association*. 75(369): 176–181.

Koenker, R. et G. Bassett Jr. 1982. « Robust Tests for Heteroscedasticity Based on Regression Quantiles ». *Econometrica*. 50(1): 43–62.

Lancaster, Kevin J. 1966. « A New Approach to Consumer Theory ». *Journal of Political Economy*. 74: 132–157.

MacDonald, Larry. 1986. « Hedonic Models of Housing: An Examination with Reference to New Housing in the Ottawa Area ». Document de travail de la Division des prix. Statistique Canada.

MacKinnon, James G., Halbert White, et Russell Davidson. 1983. « Tests for Model Specification in the Presence of Alternative Hypotheses ». *Journal of Econometrics*. 21: 53–70.

Muellbauer, John. 1974. « Household Production Theory, Quality, and the "Hedonic Technique" ». *The American Economic Review*. 64(6): 977–994.

Moulton, Brent R. 1995. « Interarea Indexes of the Cost of Shelter Using Hedonic Quality Adjustment Techniques ». *Journal of Econometrics*. 68: 181–204.

Prud'homme, Marc et Kam Yu. 1999. « Towards an Elementary Price Index for Internet Services ». Document de travail de la Division des prix. Statistique Canada.

Prud'homme, Marc et Claude Thivierge. 1999. « Spatial Price Indexes for Rental Accommodation: Progress Report ». Présenté à la réunion annuelle du Comité consultatif de la mesure des prix, 17–18 mai 1999.

Royston, J.P. 1982. « An Extension of Shapiro and Wilk's W Test for Normality to Large Samples ». *Applied Statistics*. 31(2): 115–124.

Salois, Daniel. 1987. « Comparaison des loyers entre les villes : Résultats d'une étude sur la comparaison des loyers entre les centres urbains de Montréal, Ottawa, et Toronto pour les années 1979, 1980, et 1981 ». Section centrale de recherche. Document de travail de la Division des prix. Statistique Canada.

Statistique Canada. 1992. *Guide d'utilisation des données de l'Enquête sur la population active*. Ottawa : Ministre de l'Industrie, des Sciences et de la Technologie. N° 71-528 au catalogue.

Statistique Canada. 1997. *Dictionnaire du recensement de 1996*. Ottawa : Ministre de l'Industrie. N° 92-351-XPF au catalogue.

Statistique Canada. 1998. *Méthodologie de l'Enquête sur la population active du Canada*. Ottawa : Ministre de l'Industrie. N° 71-526-XPB au catalogue.

Statistique Canada. 2000. *Guide de l'Enquête sur la population active*. Ottawa : Ministre de l'Industrie. N° 71-543-GIF au catalogue.

Annexe 1

Sources des données

Les deux sources principales de données utilisées pour l'analyse sont l'Enquête sur la population active (EPA) [août 1998 à décembre 1999] et le Recensement du Canada de 1996[11]. Après vérification et imputation des données, l'échantillon combiné comptait 14 434 ménages (annexe 2). Le questionnaire sur les loyers de l'EPA fournit les données sur les caractéristiques physiques qualitatives des logements et le Recensement fournit les données sur les caractéristiques socioéconomiques du quartier dans lequel est situé le logement.

Données de l'Enquête sur la population active

L'EPA couvre 98 % de la population canadienne. Sont exclus du champ de l'enquête les résidents des Territoires du Nord-Ouest, des réserves indiennes et des terres de la Couronne. Sont également exclus les détenus des établissements carcéraux et les membres à temps plein des Forces canadiennes, parce qu'ils sont considérés comme ne faisant pas partie du marché du travail [Guide de l'Enquête sur la population active, 2000]. L'enquête est réalisée auprès des personnes de 15 ans et plus. L'information recueillie dans le cadre de l'EPA mais non incluse dans l'analyse comprend les logements qui sont utilisés à la fois comme domicile et comme lieu d'affaires et les logements subventionnés par l'État, l'employeur ou un parent. L'EPA est basée sur un plan de sondage avec renouvellement de panel. Les ménages font partie de l'échantillon pendant six mois, et un sixième de l'échantillon est renouvelé chaque mois. Par conséquent, il existe six sous-échantillons représentatifs et un chevauchement mensuel de cinq sixièmes de l'échantillon. Le plan de sondage avec renouvellement de panel réduit le fardeau de réponse et donne de bons résultats pour l'estimation des variations mensuelles.

L'EPA comprend deux types de strates. Pour commencer, on procède à la stratification selon les régions géographiques types, qui incluent les régions économiques ainsi que les régions économiques de l'assurance-emploi. Ensuite, on forme des strates en regroupant des petits secteurs de dénombrement (SD) en régions métropolitaines de recensement (RMR) [Gambino et coll., 1998]. Une RMR est formée d'une grande région urbaine (appelée noyau urbain) comptant au moins 100 000 habitants et des régions urbaines et rurales adjacentes (appelées banlieues urbaines et rurales). Les banlieues urbaines et rurales sont des régions dont le degré d'intégration sociale et économique au noyau urbain est important [Dictionnaire du recensement de 1996 (1997)]. Le niveau d'intégration est déterminé par les flux de navetteurs qui se rendent des banlieues dans le noyau urbain et inversement.

Les renseignements sur le loyer sont recueillis lorsque le répondant participe à l'Enquête sur la population active pour la première fois, puis sont mis à jour si et quand le répondant déclare des changements par rapport aux renseignements fournis antérieurement. En ce qui concerne les modifications du prix de location, on calcule une moyenne arithmétique simple des

[11] Nous avons aussi utilisé une troisième source lors de l'agrégation des diverses composantes du logement. Pour refléter l'importance relative des éléments du logement, nous avons calculé la moyenne pondérée en nous servant des données sur les dépenses provenant de l'Enquête sur les dépenses des familles de 1996. Cette pondération par les dépenses est conforme à celle déjà utilisée dans le programme spatial.

loyers sur la période de six mois durant laquelle le répondant a fait partie de l'échantillon de l'enquête. L'échantillon total utilisé pour la présente analyse couvre la période d'août 1998 à décembre 1999. Les répondants qui entrent dans l'échantillon de l'EPA en août en sortent six mois plus tard, en janvier, et un scénario semblable est suivi pour les rotations de six mois subséquentes. Si aucun renseignement sur le loyer du répondant n'est modifié durant la période complète de six mois, seules les données du premier mois sont nécessaires pour inclure le répondant dans l'échantillon. Après correction pour tenir compte du plan de sondage avec renouvellement de l'échantillon, l'ensemble de données est complet pour l'année de référence 1999.

L'échantillon sur lequel porte l'étude est un sous-ensemble de l'échantillon de l'EPA comprenant 16 centres urbains. D'Est en Ouest, les RMR sélectionnées sont St. John's, Charlottetown (y compris Summerside), Halifax, Saint John, Fredericton, Québec, Montréal, Ottawa-Hull (partie québécoise appelée dans la suite Hull), Ottawa-Hull (partie ontarienne appelée dans la suite Ottawa), Toronto, Winnipeg, Regina, Calgary, Edmonton, Vancouver et Victoria. Tous ces centres urbains sont des RMR, sauf Charlottetown et Summerside, qui sont l'une et l'autre des agglomérations de recensement (AR).

L'EPA fournit aussi plusieurs poids que l'on peut utiliser durant l'estimation. Parmi les divers poids disponibles (poids de grappe, poids de stabilisation, poids théorique, facteur d'équilibre, sous-poids, facteur rural/urbain et poids spécial), nous choisissons le poids final, dont l'utilisation pour chaque ménage présente plusieurs avantages. Le principal est qu'on le calcule de façon qu'il représente la contribution du répondant à la population totale. Les estimations produites au moyen de ces poids concordent donc aux estimations démographiques. Le poids est calculé sous forme d'un produit de trois facteurs, à savoir un poids de sondage, qui tient compte des renseignements sur le plan de sondage (les poids de grappe et de stabilisation sont utilisés pour calculer le poids de sondage), un ajustement pour la non-réponse, qui tient compte des ménages non répondants et un facteur d'équilibre qui ajuste l'échantillon sur des chiffres de population connus. Le poids final sert aussi de rajustement pour l'erreur de couverture. Il réduit l'erreur d'échantillonnage des estimations et est le même pour tous les membres d'un ménage donné [Méthodologie de l'Enquête sur la population active du Canada, 1998].

Données de recensement

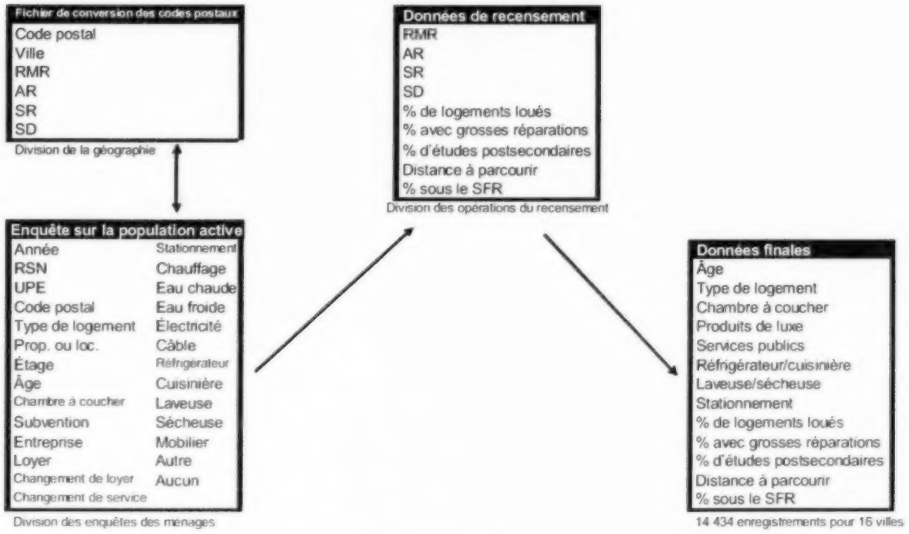
Les données de recensement utilisées ici proviennent du Recensement du Canada de 1996. Les variables de recensement pour lesquelles des données ont été obtenues sont les secteurs de dénombrement provenant de la base de données-échantillon 2B. L'échantillon à 20 % est pondéré afin de fournir des estimations pour l'ensemble de la population [Dictionnaire du recensement de 1996 (1997)]. Les secteurs de dénombrement sont les régions géographiques types les plus petites pour lesquelles des données de recensement sont publiées. Les SD sont aussi compacts que possible et leurs limites suivent des caractéristiques visibles, comme des rivières et des rues. Le nombre de logements dans un SD varie d'un minimum de 125 dans les régions rurales à un maximum de 440 dans les grandes régions urbaines [Dictionnaire du recensement de 1996 (1997)]. Certains SD sont censurés dans les données pour des raisons de confidentialité et de fiabilité statistique. Pour éviter cette perte d'information, nous avons rehaussé le niveau d'agrégation des SD censurés d'un cran en suivant trois grands critères. Seuls

les SD adjacents, les SD appartenant au même secteur de recensement (SR) et les SD capables de satisfaire aux conditions de confidentialité et de fiabilité statistiques sont regroupés. Dans certains cas, il est nécessaire de faire l'agrégation au niveau du SR pour être certain du respect de la confidentialité des données et de leur fiabilité statistique[12]. Nous supposons que les données du Recensement de 1996 sur les caractéristiques des quartiers aux niveaux du secteur de dénombrement et du secteur de recensement sont constantes jusqu'à la période couverte par l'EPA de 1999 inclusivement.

Modèle de données

Les liens illustrés à la figure A1.1 montrent l'intégration des ensembles de données décrits aux sections précédentes. Le Fichier de conversion des codes postaux (FCCP) est apparié au fichier de l'EPA en se fondant sur le code postal. Donc, les ménages individuels sont fusionnés avec leur secteur de dénombrement respectif. Cet ensemble de données combinées est alors apparié selon le SD aux données du Recensement de 1996. L'ensemble de données final, qui regroupe les données de l'EPA et celles du recensement, est celui qui est utilisé pour l'estimation des indices comparatifs entre les villes. Le tableau A1.1 donne la liste des variables endogènes et exogènes utilisées dans l'analyse, ainsi que leurs définitions.

Figure A1.1 : Modèle de données



[12] Le secteur de recensement est le niveau d'agrégation directement supérieur à celui du SD et directement inférieur à celui de l'AR ou de la RMR. La définition des secteurs de recensement est fondée sur les niveaux d'intégration sociale et économique.

Tableau A1.1

Variables des régressions

Variable dépendante :	
Loyer	Loyer mensuel total payé (Log) pour le logement.
Variables indépendantes :	
Caractéristiques physiques du logement :	
Age12	L'immeuble n'a pas plus de 5 ans (âge1) et l'immeuble a plus de 5 ans mais moins de 10 ans (âge2). L'âge de la structure est une variable discrète.
Age3	Plus de 10 ans, mais moins de 20 ans.
Age4	Plus de 20 ans, mais moins de 40 ans.
Age5	Plus de 40 ans.
DYTP1234	Cette variable fictive est une combinaison de quatre types de logement :
	DYTP1 maison unifamiliale
	DYTP2 maison jumelée
	DYTP3 maison en rangée ou en terrasse
	DYTP4 duplex non attenant
DYTP56	Cette variable est une combinaison de deux types de logement :
	DYTP5 immeuble collectif en bande comptant moins de 5 étages ou appartements
	DYTP6 tour d'habitation comptant au moins 5 étages
Floor	Numéro de l'étage où vit le répondant (variable continue).
Bedroom	Nombre déclaré de chambres à coucher dans le logement (variable continue).
Electricity	Variable fictive dont la valeur est 1 si le service est inclus dans le loyer et 0 autrement.
Luxuries	Cette variable fictive est une variable composite des biens et services inclus dans le loyer mensuel : câblodistribution, mobilier ou autres gros appareils ménagers.
Utilities	Cette variable fictive est une combinaison des services publics inclus dans le loyer mensuel : chauffage et eau chaude.
Fridge/Stove	Combinaison de réfrigérateur et cuisinière.
Washdry	Combinaison de laveuse et sécheuse.
Parking	Variable fictive dont la valeur est égale à 1 si le stationnement est inclus dans le loyer mensuel.
Caractéristiques du quartier (secteur de dénombrement) :	
Var4	Pourcentage de logements privés loués dans le secteur de dénombrement.
Var7	Pourcentage de logements privés ayant besoin de grosses réparations. Par grosses réparations, on entend la réparation d'une installation de plomberie ou d'une installation électrique défectueuse, des réparations structurelles aux murs, aux planchers ou aux plafonds, etc.
Var11	Pourcentage de personnes qui ont fait des études postsecondaires. Par études postsecondaires, on entend l'obtention d'un certificat, d'un diplôme ou d'un grade décerné par un collège communautaire ou une université.
Var41	Distance moyenne parcourue pour se rendre au travail (km). La variable de distance moyenne parcourue est une variable discrète, mesurée par tranche de 5 km jusqu'à 30 km ou plus.
Var43	Pourcentage de personnes vivant sous le seuil de faible revenu.

Annexe 2

Description de l'échantillon

Tableau A2.1

Emplacement	N
St. John's	235
Charlottetown	611
Halifax	819
Saint John	337
Fredericton	225
Québec	793
Montréal	2 496
Hull	536
Ottawa	770
Toronto	2 319
Winnipeg	1 424
Regina	511
Calgary	656
Edmonton	762
Vancouver	1 400
Victoria	540
Total	**14 434**

Tableau A2.2

Pourcentage des divers types de logements selon la région métropolitaine de recensement*

	Maison unifamiliale	Maison jumelée	Maison en rangée ou en terrasse	Duplex	Immeuble collectif en bande	Tour d'habitation
St. John's	10,6	12,3	9,8	37,9	27,2	2,1
Charlottetown	11,6	8,3	3,8	4,3	70,7	0,0
Halifax	7,6	3,8	2,0	7,0	54,0	23,4
Saint John	6,5	7,4	1,2	6,5	66,2	10,7
Fredericton	8,9	8,0	2,7	5,3	67,1	6,2
Québec	3,7	3,8	3,2	5,8	76,8	6,2
Montréal	3,3	1,0	5,3	7,1	68,1	15,0
Hull	11,0	8,8	3,7	8,2	60,6	6,5
Ottawa	7,1	4,5	13,1	4,5	31,0	39,5
Toronto	7,8	4,7	5,5	1,6	21,2	58,4
Winnipeg	9,8	3,3	6,0	2,5	45,0	32,4
Regina	24,1	4,9	13,3	1,0	47,9	8,4
Calgary	20,7	6,3	13,0	8,5	33,4	18,0
Edmonton	17,8	2,1	13,1	2,5	42,9	18,8
Vancouver	17,9	3,6	4,4	7,6	42,6	21,2
Victoria	20,4	5,2	5,0	7,0	55,4	6,3

* Les données ayant été arrondies, la somme des chiffres pour chaque RMR pourrait ne pas être égale à 100 %.

Tableau A2.3
Pourcentage de structures dans la fourchette d'âge (en années) selon la région métropolitaine de recensement*

	(< 5 ans)	(5 < âge < 10)	(10 < âge < 20)	(20 < âge < 40)	(> 40 ans)
St. John's	4,3	12,3	26,0	34,0	23,4
Charlottetown	3,8	11,0	23,6	23,4	38,3
Halifax	5,4	13,8	17,9	42,1	20,8
Saint John	0,0	3,0	7,1	29,7	60,2
Fredericton	12,0	6,7	16,4	33,3	31,6
Québec	1,6	11,6	16,1	40,6	30,0
Montréal	1,0	4,7	12,3	45,8	36,3
Hull	4,5	12,7	14,4	40,9	27,6
Ottawa	2,1	5,7	12,3	53,5	26,4
Toronto	1,7	4,0	17,0	52,6	24,7
Winnipeg	0,5	2,3	16,6	53,4	27,2
Regina	1,6	2,2	16,4	47,6	32,3
Calgary	1,4	1,8	17,4	60,7	18,8
Edmonton	2,4	1,8	16,3	62,5	17,1
Vancouver	5,6	6,9	17,1	48,3	22,1
Victoria	2,8	5,0	10,7	52,2	29,3

* Les données ayant été arrondies, la somme des chiffres pour chaque RMR pourrait ne pas être égale à 100 %.

Tableau A2.4
Pourcentage de logements selon le nombre de chambres à coucher et selon la région métropolitaine de recensement*

	Studio	Une	Deux	Trois	Quatre
St. John's	1,3	23,4	47,2	24,3	3,8
Charlottetown	2,3	29,0	48,4	16,4	3,9
Halifax	5,3	36,6	40,4	15,4	2,3
Saint John	4,7	23,4	43,0	26,4	2,4
Fredericton	8,4	20,9	51,6	15,6	3,6
Québec	4,5	31,1	50,7	11,9	1,8
Montréal	6,5	32,1	41,2	18,8	1,4
Hull	4,7	26,5	44,4	21,8	2,6
Ottawa	5,6	36,1	33,1	21,4	3,8
Toronto	5,5	38,8	37,0	16,1	2,6
Winnipeg	6,0	44,2	37,9	10,3	1,8
Regina	4,3	31,1	42,5	19,6	2,5
Calgary	1,8	31,7	38,9	19,5	8,1
Edmonton	7,9	34,5	30,3	22,3	5,0
Vancouver	8,1	41,7	30,4	14,2	5,6
Victoria	7,0	40,0	37,6	11,1	4,3

* Les données ayant été arrondies, la somme des chiffres pour chaque RMR pourrait ne pas être égale à 100 %.

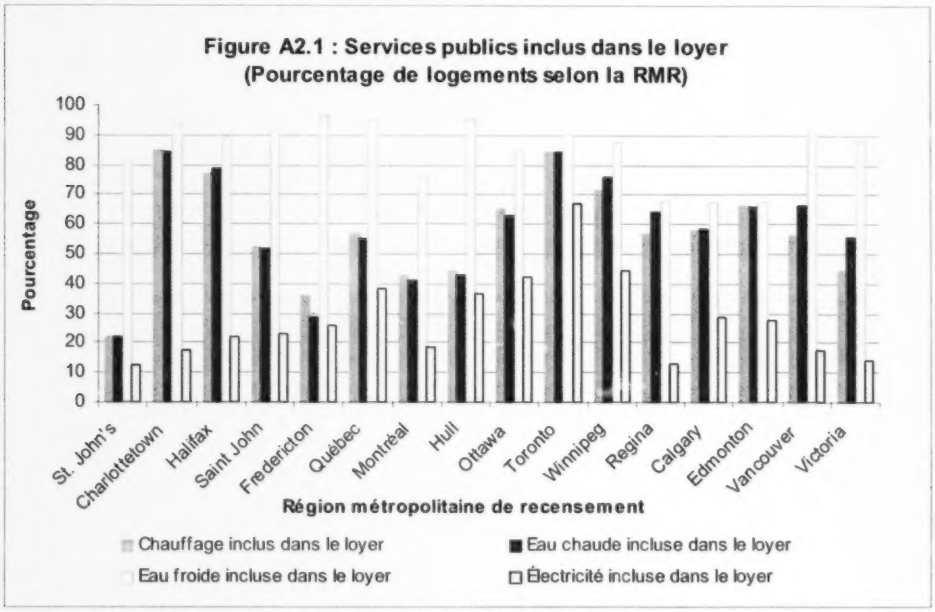


Figure A2.1 : Services publics inclus dans le loyer
(Pourcentage de logements selon la RMR)
Pourcentage
100
90
80
70
60
50
40
30
20
10
0
St. John's
Charlottetown
Halifax
Saint John
Fredericton
Québec
Montréal
Hull
Ottawa
Toronto
Winnipeg
Regina
Calgary
Edmonton
Vancouver
Victoria
Région métropolitaine de recensement
Chauffage inclus dans le loyer
Eau chaude incluse dans le loyer
Eau froide incluse dans le loyer
Électricité incluse dans le loyer

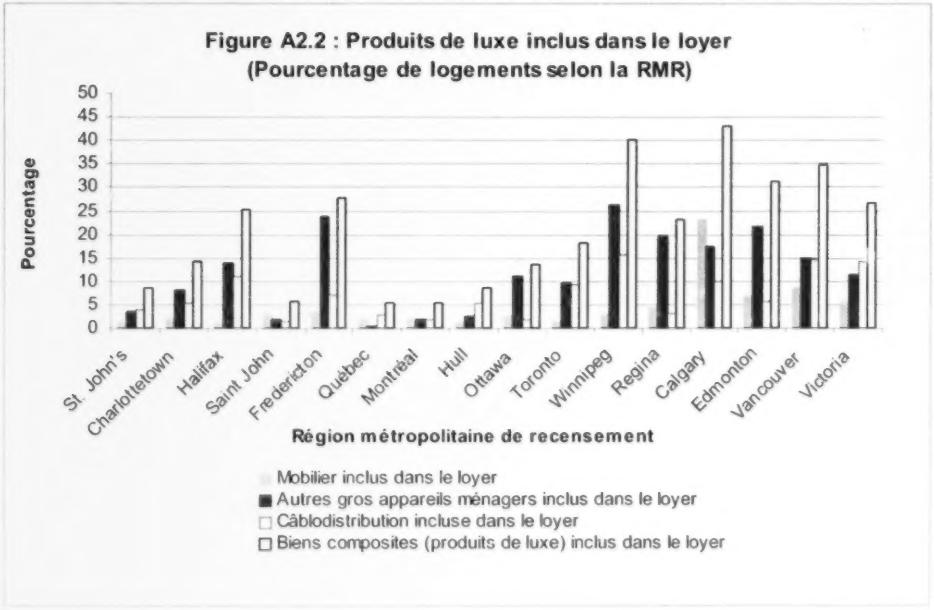


Figure A2.2 : Produits de luxe inclus dans le loyer
(Pourcentage de logements selon la RMR)
Pourcentage
50
45
40
35
30
25
20
15
10
5
0
St. John's
Charlottetown
Halifax
Saint John
Fredericton
Québec
Montréal
Hull
Ottawa
Toronto
Winnipeg
Regina
Calgary
Edmonton
Vancouver
Victoria
Région métropolitaine de recensement
Mobilier inclus dans le loyer
Autres gros appareils ménagers inclus dans le loyer
Câblodistribution incluse dans le loyer
Biens composites (produits de luxe) inclus dans le loyer

Figure A2.3 : Laveuse et sécheuse incluses dans le loyer (Pourcentage de logements selon la RMR)

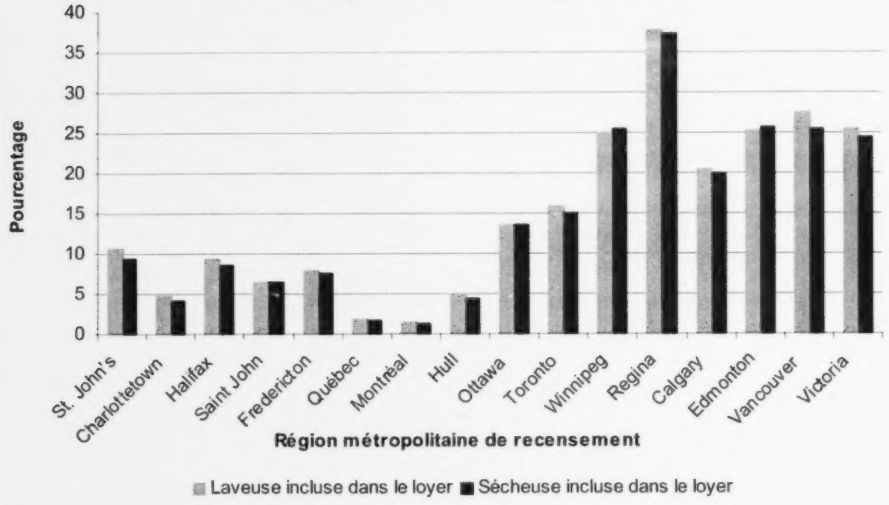
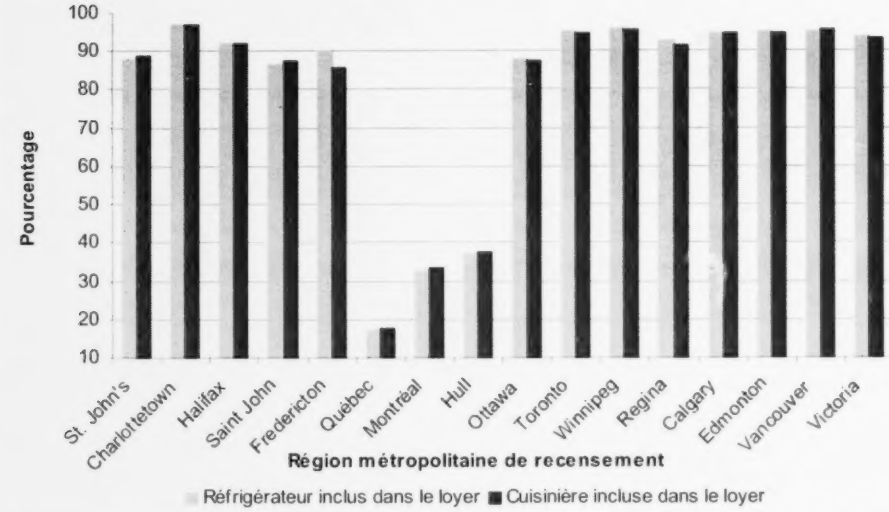


Figure A2.4 : Réfrigérateur et cuisinière inclus dans le loyer (Pourcentage de logements selon la RMR)



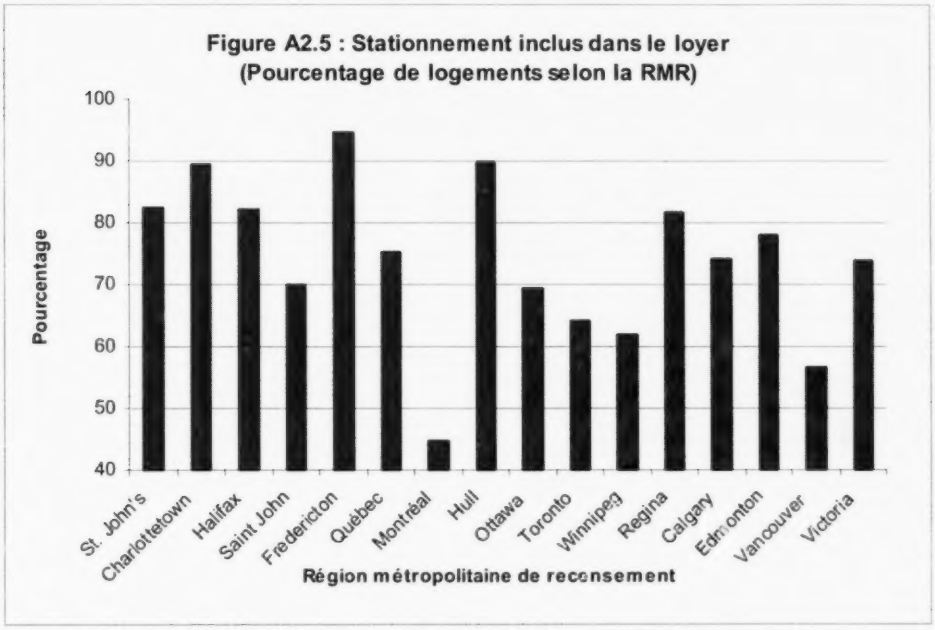


Tableau A2.5

Moyennes des caractéristiques socioéconomiques selon la région métropolitaine de recensement

	Pourcentage de logements loués	Pourcentage avec grosses réparations	Pourcentage avec études postsecondaires	Distance à parcourir (km)	Pourcentage de logements sous le seuil de faible revenu
St. John's	43,1	6,1	54,9	6,1	17,8
Charlottetown	53,9	7,3	50,5	5,2	19,6
Halifax	67,2	6,3	60,3	5,7	23,7
Saint John	62,8	9,9	43,1	5,6	27,3
Fredericton	62,2	10,0	60,1	5,8	20,8
Québec	64,9	5,8	46,3	5,4	28,5
Montréal	69,8	7,9	46,8	5,0	31,3
Hull	54,6	8,4	44,3	5,0	26,5
Ottawa	64,9	8,6	60,1	5,8	24,3
Toronto	68,8	9,3	52,3	5,9	27,8
Winnipeg	66,0	8,3	47,1	4,5	29,3
Regina	54,1	8,2	49,4	5,0	22,8
Calgary	59,5	8,2	59,0	6,1	22,8
Edmonton	61,5	6,4	53,6	6,9	25,4
Vancouver	58,7	7,8	60,6	5,8	22,4
Victoria	55,1	6,3	60,4	5,8	16,4

Tableau A2.6

Médianes des caractéristiques socioéconomiques selon la région métropolitaine de recensement

	Pourcentage de logements loués	Pourcentage avec grosses réparations	Pourcentage avec études postsecondaires	Distance à parcourir (km)	Pourcentage de logements sous le seuil de faible revenu
St. John's	40,8	4,5	56,2	5,7	15,4
Charlottetown	51,2	6,0	50,4	4,5	17,6
Halifax	74,7	5,5	58,7	4,9	19,0
Saint John	70,8	8,3	42,2	5,6	22,2
Fredericton	68,7	8,4	62,6	6,1	21,4
Québec	72,9	5,4	43,2	5,0	25,7
Montréal	72,6	7,2	44,3	4,6	30,8
Hull	54,0	7,0	42,9	4,7	24,5
Ottawa	66,7	8,8	59,3	5,1	21,6
Toronto	73,4	8,3	50,3	5,5	25,0
Winnipeg	69,6	6,3	47,1	4,3	26,1
Regina	50,7	7,9	49,5	4,5	22,2
Calgary	62,1	7,8	59,4	5,3	20,8
Edmonton	61,5	5,6	51,4	6,4	22,0
Vancouver	59,4	7,1	59,3	5,4	22,2
Victoria	58,7	5,3	61,7	5,5	10,8

Annexe 3

Spécification et estimation du modèle

Modèle statistique

Nous utilisons deux tests pour explorer les spécifications du modèle, à savoir la transformation de Box-Cox et le test PE de MacKinnon, White et Davidson [1983]. Pour commencer, nous utilisons la transformation de Box-Cox de façon itérative pour tester diverses spécifications qui retiennent leur aspect pratique du point de vue analytique [Godfrey, McAleer et McKenzie, 1988]. Donc, nous réalisons une transformation de Box-Cox sur la variable dépendante de loyer pour chaque emplacement.

$y_k(\lambda) = \alpha + \beta_j x_{jk} + \varepsilon_k$, où $y_k(\lambda)$ dénote la transformation. (A3.1)

$$\begin{aligned} y_k(\lambda) &= \left(y_k^{\lambda} - 1\right)/\lambda, \quad \lambda \neq 0, \quad y_k \geq 0 \\ &= \log y_k \quad , \quad \lambda = 0, \quad y_k > 0 \end{aligned} \tag{A3.2}$$

L'itération de λ va de -1 à 3 par incréments de 0,1. Nous utilisons le R^2 corrigé et le critère d'information d'Akaike (AIC pour *Akaike's information criterion*) comme mesure de la qualité de l'ajustement en vue de comparer les diverses spécifications. Le R^2 corrigé atteint sa valeur la plus élevée, soit 0,4842 quand λ=0,1, tandis que la valeur correspondante de l'AIC est de 3687. La spécification λ=−0,2 est celle pour laquelle le R^2 corrigé est le plus faible, mais l'AIC, qui est de 236, s'approche de zéro (tableau A3.1). Pour appuyer davantage le choix de la spécification semi-logarithmique, nous avons utilisé le test PE élaboré par MacKinnon, White et Davidson. Ce test est non emboîté et l'hypothèse de normalité n'est pas nécessaire (contrairement au cas de la transformation de Box-Cox). En principe, le test PE ne présente pas de propriété d'optimalité, et on peut s'attendre à ce qu'il soit moins puissant que les tests de Wald, du multiplicateur de Lagrange (LM) et du rapport des vraisemblances (LR). Cependant, selon certains, malgré ces lacunes théoriques, le test PE demeure un outil puissant, facile à utiliser [MacKinnon, White et Davidson, 1983]. Le terme $\log \hat{y}_k$ est la valeur prévue de $\log y_k$ pour la spécification semi-logarithmique, et $\log \tilde{y}_k$ représente le logarithme de la valeur prévue de y_k dans les équations suivantes :

$$\log y_k = \alpha + \beta_j x_{jk} + \gamma_0 \left[\tilde{y}_k - e^{\left(\log \hat{y}_k\right)} \right] + \varepsilon_k \tag{A3.3}$$

et

$$y_k = \alpha + \beta_j x_{jk} + \gamma_1 \left[\log \hat{y}_k - \log \tilde{y}_k \right] + \varepsilon_k \tag{A3.4}$$

L'hypothèse nulle de non-rejet de la spécification semi-logarithmique s'écrit $\gamma_0=0$ et l'hypothèse nulle de non-rejet de la spécification linéaire s'écrit $\gamma_1=0$. Le test indique que la spécification semi-logarithmique est celle qui est la mieux adaptée aux données analysées (tableau A3.2).

Tableau A3.1

Test de spécification de Box-Cox

Lambda	R^2	R^2 corrigé	AIC
-0,5	0,4811	0,4662	-3 146
-0,4	0,4868	0,4721	-2 026
-0,3	0,4915	0,4769	-898
-0,2	0,4950	0,4805	236
-0,1	0,4974	0,4830	1 379
0	0,4984	0,4841	2 530
0,1	0,4985	0,4842	3 687
0,2	0,4971	0,4828	4 853
0,3	0,4944	0,4801	6 028

Il s'agit de la valeur moyenne des statistiques de qualité de l'ajustement pour l'ensemble des emplacements.

Tableau A3.2

Test PE de MacKinnon, White et Davidson

IPC	Emplacement	Statistique t
1	St. John's	-2,34
2	Charlottetown	0,04
3	Halifax	-3,34
4	Saint John	0,26
5	Fredericton	-5,39
6	Québec	0,74
7	Montréal	-8,17
8	Hull	-0,83
9	Ottawa	0,45
10	Toronto	-1,90
11	Winnipeg	4,19
12	Regina	-3,01
13	Calgary	-0,43
14	Edmonton	4,18
15	Vancouver	-1,44
16	Victoria	1,35

Pour ne pas rejeter l'hypothèse nulle de spécification semi-logarithmique, il faut que le paramètre soit égal à zéro ou que $-1,95 < t < 1,95$ au niveau de signification de 5 %.

Nous avons calculé la régression du logarithme naturel du prix de location sur les attributs physiques et socioéconomiques selon l'emplacement. Le modèle semi-logarithmique des régressions distinctes peut s'exprimer sous la forme :

$$Y_{vk} = \alpha_v + \sum_{j=1}^{J} \beta_{vj} X_{vjk} + \varepsilon_{vk} \qquad \text{(A3.5)}$$

$$E(\varepsilon_{vk}) = 0, \qquad \text{(A3.6)}$$

$$E(X_{vjk}\ \varepsilon_{vk}) = 0, \qquad \text{(A3.7)}$$

$$E(\varepsilon^2_{vk}) = \sigma^2_{vk} \qquad \text{(A3.8)}$$

où Y_{vk} est le logarithme naturel du loyer mensuel payé pour le k^e logement à l'emplacement v. α_v est la coordonnée à l'origine de l'emplacement v. β_{vj} est le vecteur de coefficients pour les j caractéristiques qualitatives du logement, et X_{vjk} est la matrice de tous les attributs (physiques et socioéconomiques). ε_{vk} représente le terme d'erreur pour le k^e logement à l'emplacement v.

Méthodes statistiques

À la présente section, nous discutons des difficultés que pose l'estimation, puis nous examinons brièvement les méthodes utilisées pour les résoudre. Le calcul des indices hédoniques se fait en deux étapes. Premièrement, il faut obtenir les coefficients des caractéristiques qualitatives au moyen d'une régression du prix de location sur les caractéristiques qualitatives. À la deuxième étape, on utilise les estimations ainsi obtenues pour construire un indice de prix spatial de Fisher, qui est une moyenne géométrique des indices de prix de Laspeyres et de Paasche.

Avant de pouvoir exécuter les régressions, il faut résoudre les problèmes posés par la multicollinéarité et les variables omises. Comme nous avons observé une multicollinéarité pour certaines variables, nous avons exécuté des tests de corrélation pour toutes les variables incluses dans les régressions. Il existe une corrélation entre les variables de chauffage et d'eau chaude, les variables de réfrigérateur et de cuisinière et celles de laveuse et de sécheuse. Pour régler le problème de multicollinéarité, nous avons créé des produits composites (annexe 2). Cette approche est justifiée en raison de la corrélation presque parfaite entre les variables susmentionnées. Étant donné la nature complémentaire des biens, la façon de procéder évidente consiste à regrouper les flux de services émanant de ces biens. Nous avons créé une variable composite pour la laveuse et la sécheuse, en fixant la valeur de « washdry » à un si une laveuse et une sécheuse sont incluses dans le loyer, et à zéro autrement. Nous faisons le même calcul pour le réfrigérateur et la cuisinière. La variable fictive pour les services publics est égale à un si le chauffage et l'eau chaude sont inclus dans le loyer et est nulle autrement. La variable d'électricité est incluse dans la régression séparément. Le calcul pour les produits de luxe (comme la câblodistribution, le mobilier et d'autres gros appareils ménagers) est différent. La variable fictive de produits de luxe est égale à un si le loyer inclut la câblodistribution, le mobilier ou d'autres gros appareils ménagers et est nulle autrement. L'âge de la structure, le type de logement et le pourcentage de logements privés loués sont des exemples de variables présentant des corrélations problématiques avec d'autres variables. Rien n'a toutefois été fait pour résoudre le problème de multicollinéarité de ces variables parce que les coefficients de régression n'affichaient pas systématiquement un signe ou un ordre de grandeur invraisemblable et que l'ordre de grandeur des erreurs-types ne dépassait pas les attentes. Deux problèmes, à

savoir la limitation des données et les problèmes de collinéarité, sont la cause des variables omises. Une liste partielle de ces dernières inclut l'existence d'une piscine, la climatisation centrale, le nombre de pieds carrés du logement, le nombre de pièces dans le logement, la valeur du terrain et, éventuellement, certaines variables environnementales, comme un indice de pollution de l'air ou un indice de vue ou d'horizon. Il convient de souligner qui existera toujours des variables auxquelles le consommateur accorde de la valeur mais qu'il est impossible d'observer et, par conséquent, d'intégrer dans le modèle hédonique.

Nous avons exécuté une régression distincte pour chaque région pour laquelle un indice est produit (c.-à-d. chaque centre urbain). La méthode des moindres carrés ordinaires (MCO) n'est pas appropriée ici à cause de la nature non sphérique des erreurs. Étant donné la nature de notre étude transversale, nous avons supposé qu'il pourrait exister une hétéroscédasticité. Nous avons exécuté des tests de spécification préliminaire de White pour analyser les spécifications et l'hétéroscédasticité. Les résultats sont contradictoires en ce qui concerne le rejet de l'hypothèse nulle d'homoscédasticité pour tous les emplacements, mais il faudra réaliser d'autres tests car ceux de White sont généraux.

Les propriétés asymptotiques du grand échantillon nous ont incités à choisir un test du multiplicateur de Breusch-Pagan-Godfrey (BPG) pour rechercher toute présence significative d'une hétéroscédasticité. Le test BPG est un test du multiplicateur de Lagrange (LM) de l'hypothèse nulle d'homoscédasticité. La statistique LM est calculée en divisant la somme expliquée des carrés de la régression de e^2_{vk} sur z_{vk} par les MCO et un terme de coordonnée à l'origine par $2s^4$, où $s^2 = \frac{1}{K}\sum_k e^2_{vk}$. Ce test est assez sensible à l'hypothèse de normalité.

Nous avons fait un test de Shapiro-Wilk (SW) de grand échantillon pour évaluer la normalité selon l'emplacement. Les résultats de ce test ont montré que l'hypothèse de normalité n'est pas justifiée. Royston [1982] a introduit l'extension du test SW de grand échantillon suivante :

$$W_{vk} = 1 - \left[(z_{vk})\sigma_y + \mu_y\right]^{V_\tau}, \qquad 0 < W_{vk} \leq 1 \qquad \text{(A3.9)}$$

où z_{vk} est une variable normale centrée réduite, et μ_y et σ_y sont la moyenne et l'écart-type de y. Les paramètres τ, μ_y et σ_y sont des fonctions de vk obtenues d'après les résultats de simulations (voir Royston [1982] pour des précisions). Un écart par rapport à la normalité est indiqué par une faible valeur de W. Le test de Shapiro-Wilk donne lieu au rejet de l'hypothèse que l'ensemble de données agrégées obéit à une loi normale au niveau de signification de 0,01 %. Lorsque les données sont réparties en sous-échantillons selon l'emplacement, le test indique que l'hypothèse de normalité n'est pas justifiée non plus au niveau de signification de 0,01 %, et ce pour tous les emplacements (tableau A3.3).

Tableau A3.3

Test de Shapiro-Wilk pour la normalité

IPC	Emplacement	Stat. p
1	St. John's	0,0001
2	Charlottetown	0,0001
3	Halifax	0,0001
4	Saint John	0,0001
5	Fredericton	0,0015
6	Québec	0,0001
7	Montréal	0,01
8	Hull	0,0001
9	Ottawa	0,0001
10	Toronto	0,01
11	Winnipeg	0,0001
12	Regina	0,0001
13	Calgary	0,0001
14	Edmonton	0,0001
15	Vancouver	0,0001
16	Victoria	0,0001

Si la valeur de la statistique p est égale ou inférieure à 0,01, l'hypothèse nulle d'une loi normale est rejetée (niveau de signification de 1 %).

Koenker et Bassett [1982] ont proposé une modification du calcul de la statistique LM pour le test BPG, quand l'hypothèse de normalité n'est pas appropriée. La statistique LM proposée est de la forme :

$$LM = \left(\frac{1}{H}\right)\left(u - \bar{u}\,i\right)' Z\left(Z'Z\right)^{-1} Z' \left(u - \bar{u}\,i\right) \qquad \text{(A3.10)}$$

où u est $(e^2_1, e^2_2, \ldots, e^2_K)$, $\bar{u} = \dfrac{e'e}{K}$ et i est une colonne K fois un de valeurs 1. Le calcul est fondé sur un estimateur plus robuste de la variation de ${\varepsilon_{ik}}^2$. Si elle obéit à la loi normale, la variance de ${\varepsilon_{ik}}^2$ n'est pas nécessairement égale à $2\sigma^4$. Donc, la modification est $H = \dfrac{1}{K}\sum_k \left(e^2_{ik} - \dfrac{e'e}{K}\right)^2$.

En l'absence de normalité, le calcul de Koenker et Bassett fournit un test plus puissant. Cependant, sous l'hypothèse de normalité, la statistique modifiée obéira asymptotiquement à la même loi que la statistique LM de Breusch-Pagan. La statistique du multiplicateur de Lagrange suit asymptotiquement la même loi que le chi carré dont le nombre de degrés de liberté est égal au nombre de variables dans la régression. Les tests de Koenker et Bassett sont tous statistiquement significatifs au niveau de signification de 0,05 % (sauf pour la ville de Québec, tableau A3.4); donc, nous rejetons l'hypothèse nulle d'homoscédasticité au profit de l'hypothèse alternative d'hétéroscédasticité.

Tableau A3.4

Tests BPG pour l'hétéroscédasticité

IPC	Emplacement	Statistique LM
1	St. John's	48,9
2	Charlottetown	37,1
3	Halifax	82,2
4	Saint John	34,0
5	Fredericton	47,3
6	Québec	13,7
7	Montréal	83,9
8	Hull	44,2
9	Ottawa	32,5
10	Toronto	48,7
11	Winnipeg	88,6
12	Regina	130,0
13	Calgary	85,4
14	Edmonton	115,7
15	Vancouver	39,6
16	Victoria	56,4

Pour DF=15, la statistique du chi carré=26,3 au niveau de signification de 5 %. Si la valeur de la statistique LM est supérieure à celle de la statistique du chi carré, l'hypothèse nulle d'homoscédasticité est rejetée.

Nous employons une méthode d'estimation par les moindres carrés généralisés réalisables (MCGR) en deux étapes pour corriger pour l'hétéroscédasticité, c'est-à-dire la variation du terme d'erreur aléatoire à l'intérieur des classes d'attributs. Par classe d'attributs, nous entendons la mesure discrète ou binaire de l'âge de la structure, du type de logement ou des variables de biens et services. Nous utilisons les termes d'erreur des MCO pour obtenir une estimation convergente de Ω par remplacement des paramètres réels, mais inconnus, de covariance des erreurs. Cette technique est asymptotiquement équivalente aux moindres carrés généralisés (MCG) qui convergent en loi vers l'estimateur du maximum de vraisemblance et, par conséquent, les estimations sont sans biais, convergentes et efficaces comparativement aux moindres carrés ordinaires [Jobson et Fuller, 1980]. Une procédure générale proposée par Amemiya [1985] consiste à utiliser les résidus des régressions par les moindres carrés ordinaires comme variable endogène pour estimer les paramètres θ :

$$Var\left(\varepsilon_{ik}^2\right) = \sigma_{ik}^2\left(\theta' Z_{ik}\right) \qquad \text{(A3.11)}$$

Le Z est un ensemble de variables indépendantes qui, ici, coïncident avec X. Les θ sont des estimations convergentes des paramètres inconnus de Ω, avec l'information qui suit provenant de la régression :

$$\sigma_{ik}^2 = \theta' Z \text{ ou, de façon équivalente,} \qquad \text{(A3.12)}$$

$$e_{ik}^2 = \theta' Z_{ik} + \upsilon_{ik} \qquad \text{(A3.13)}$$

Nous pouvons utiliser la valeur prévue de e_{vk}^2 comme estimation de σ_{vk}^2, puisqu'elles sont asymptotiquement équivalentes. Donc, l'estimateur employé fournit une estimation en deux étapes de β.

$$\hat{\beta}_{vj}^{*} = \left[\sum_{k=0}^{K} \frac{1}{\sigma_{vk}^2} X'_{vjk} X_{vjk}\right]^{-1} \left[\sum_{k=0}^{K} \frac{1}{\sigma_{vk}^2} X'_{vjk} Y_{vk}\right] \qquad \text{(A3.14)}$$

Cet estimateur est équivalent à l'estimateur par les moindres carrés pondérés si $w_{vk} = 1/\sigma_{vk}^2$.

$$\hat{\beta}_{vj}^{w} = \left[\sum_{k=0}^{K} w_{vk} X'_{vjk} X_{vjk}\right]^{-1} \left[\sum_{k=0}^{K} w_{vk} X'_{vjk} Y_{vk}\right] \qquad \text{(A3.15)}$$

L'hypothèse implicite que les termes d'erreur suivent la loi normale est problématique si la distribution n'est pas normale. Carroll et Ruppert [1982a] proposent une méthode qui est robuste aux valeurs aberrantes et aux lois non normales. Ils démontrent que les estimations par les moindres carrés généralisés pondérés sont asymptotiquement tout aussi bonnes que si les pondérations étaient effectivement connues.

La deuxième étape de la production des indices interrégionaux consiste à appliquer les coefficients ainsi estimés à la construction d'un indice de prix spatial de Fisher-Törnqvist (FT) qui est une moyenne géométrique des indices de prix de Laspeyres (L) et de Paasche (P). En utilisant une moyenne géométrique des indices de Laspeyres et de Paasche, on utilise la moyenne des *j* caractéristiques de la région de référence ainsi que celle des *j* caractéristiques de la région comparée. Toutefois, le problème de la non-transivité de l'indice de Fisher-Törnqvist persiste.

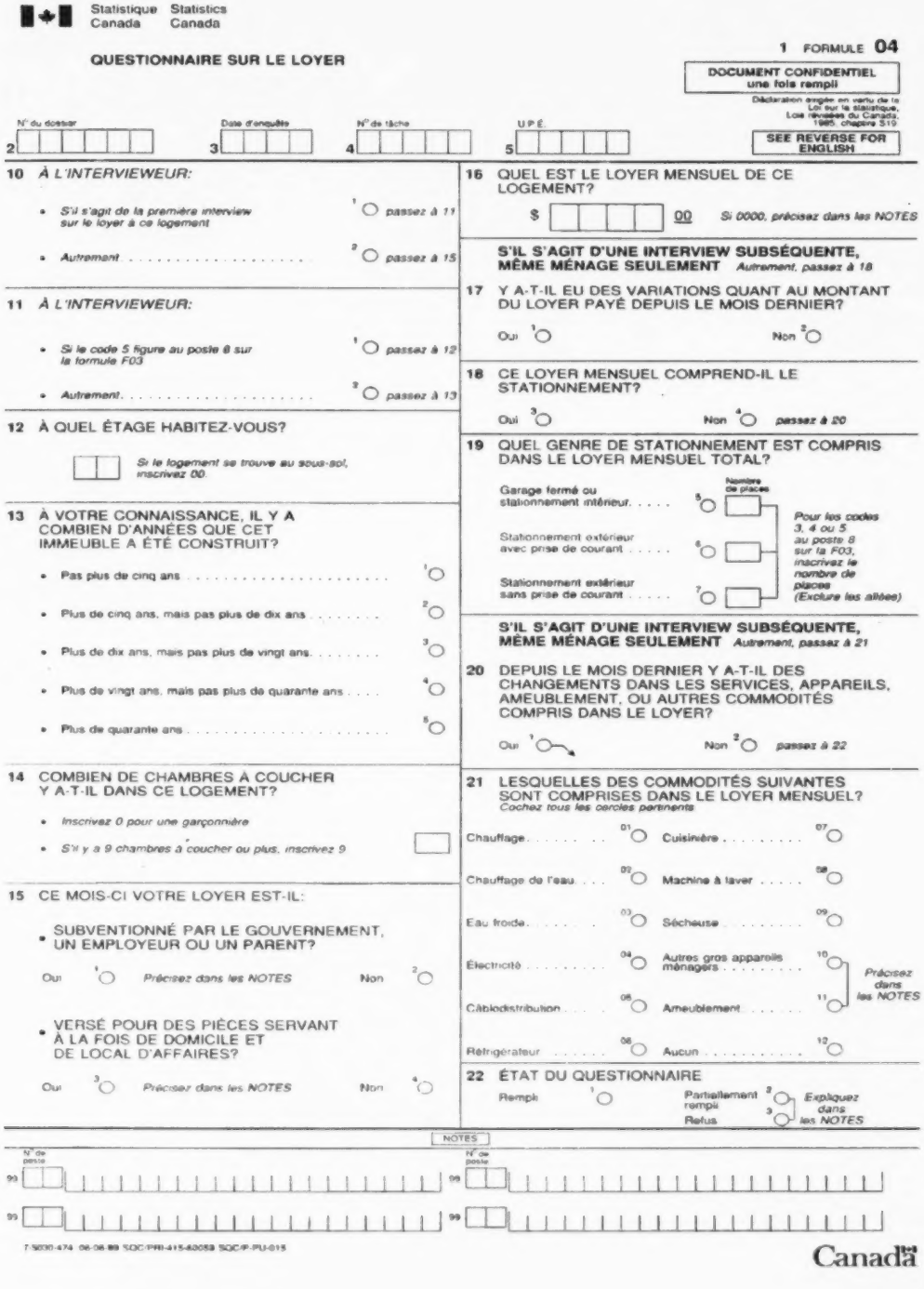
Les propriétés désirables minimales auxquelles tout indice temporel construit raisonnablement devrait satisfaire sont les tests d'identité, de proportionnalité et de changement d'unité. Ces propriétés tiennent également pour les comparaisons spatiales; cependant, il faut remplacer les estimations de prix par les prix implicites des caractéristiques qualitatives. Le test d'identité est satisfait si l'on compare une région à elle-même et que l'indice ne varie pas. Le test de proportionnalité est satisfait si, quand tous les prix évoluent proportionnellement, l'indice en fait de même. Enfin, le test de changement d'unité est satisfait si les estimations des prix sont « constantes sous toute modification des unités monétaires ou physiques dans lesquelles les prix individuels sont mesurés » [Allen, 1975]. Les indices de Laspeyres, de Paasche et de Fisher satisfont tous à ces tests; cependant, les indices de Laspeyres et de Paasche ne satisfont pas les tests pour deux propriétés de transitivité, c'est-à-dire l'inversion du temps et la circularité. L'indice de Fisher, quant à lui, passe le test d'inversion du temps ($P_{st} = 1/P_{ts}$, où *s* n'est pas égal à *t*, et *s* et *t* = 0, 1, 2...). Toutefois, il ne passe pas le test circulaire ($P_{ts}*P_{st} = P_{tt}$, où *s* n'est pas égal à *t*, et *s* et *t* = 0, 1, 2...). Si le test circulaire est satisfait, l'évolution des prix au cours du temps n'a pas d'importance et, d'un point de vue économique, il s'agit d'une hypothèse très contraignante.

Tableau A3.5

Indices spatiaux des prix du logement par paires de villes

Numéro de la paire	Paire de villes	LASPEYRES	PAASCHE	FISHER
1	St. John's et Halifax (100)	98,5	92,0	95,2
2	Charlottetown et Halifax (100)	87,4	88,7	88,0
3	Saint John et Halifax (100)	82,1	82,6	82,4
4	Ottawa et Halifax (100)	127,7	123,3	125,5
5	Toronto et Montréal (100)	163,1	160,0	161,6
6	Ottawa et Toronto (100)	81,0	80,5	80,7
7	Toronto et Winnipeg (100)	162,6	174,2	168,3
8	Regina et Winnipeg (100)	92,2	93,6	92,9
9	Edmonton et Winnipeg (100)	99,6	103,9	101,7
10	Edmonton et Vancouver (100)	69,1	74,3	71,6
11	Fredericton et Halifax (100)	101,4	103,8	102,6
12	Québec et Montréal (100)	90,6	90,6	90,6
13	Calgary et Edmonton (100)	115,5	118,4	116,9
14	Victoria et Vancouver (100)	93,4	90,2	91,8

Annexe 4



Statistique Canada Statistics Canada

QUESTIONNAIRE SUR LE LOYER

1 FORMULE 04

DOCUMENT CONFIDENTIEL une fois rempli

Déclaration exigée en vertu de la Loi sur la statistique, Lois révisées du Canada, 1985, chapitre S-19

SEE REVERSE FOR ENGLISH

2 N° du dossier | 3 Date d'enquête | 4 N° de tâche | 5 U.P.É.

10 *À L'INTERVIEWEUR:*

- *S'il s'agit de la première interview sur le loyer à ce logement* 1 ○ *passez à 11*
- *Autrement* . . . . . 2 ○ *passez à 15*

11 *À L'INTERVIEWEUR:*

- *Si le code 5 figure au poste 8 sur la formule F03* 1 ○ *passez à 12*
- *Autrement* . . . . . 2 ○ *passez à 13*

12 À QUEL ÉTAGE HABITEZ-VOUS?

☐☐ *Si le logement se trouve au sous-sol, inscrivez 00.*

13 À VOTRE CONNAISSANCE, IL Y A COMBIEN D'ANNÉES QUE CET IMMEUBLE A ÉTÉ CONSTRUIT?

- Pas plus de cinq ans . . . . . 1 ○
- Plus de cinq ans, mais pas plus de dix ans . . . . . 2 ○
- Plus de dix ans, mais pas plus de vingt ans. . . . . 3 ○
- Plus de vingt ans, mais pas plus de quarante ans . . . . 4 ○
- Plus de quarante ans . . . . . 5 ○

14 COMBIEN DE CHAMBRES À COUCHER Y A-T-IL DANS CE LOGEMENT?

- *Inscrivez 0 pour une garçonnière*
- *S'il y a 9 chambres à coucher ou plus, inscrivez 9* ☐

15 CE MOIS-CI VOTRE LOYER EST-IL:

- SUBVENTIONNÉ PAR LE GOUVERNEMENT, UN EMPLOYEUR OU UN PARENT?

  Oui 1 ○ *Précisez dans les NOTES* Non 2 ○

- VERSÉ POUR DES PIÈCES SERVANT À LA FOIS DE DOMICILE ET DE LOCAL D'AFFAIRES?

  Oui 3 ○ *Précisez dans les NOTES* Non 4 ○

16 QUEL EST LE LOYER MENSUEL DE CE LOGEMENT?

$ ☐☐☐☐ 00 *Si 0000, précisez dans les NOTES*

S'IL S'AGIT D'UNE INTERVIEW SUBSÉQUENTE, MÊME MÉNAGE SEULEMENT *Autrement, passez à 18*

17 Y A-T-IL EU DES VARIATIONS QUANT AU MONTANT DU LOYER PAYÉ DEPUIS LE MOIS DERNIER?

Oui 1 ○ Non 2 ○

18 CE LOYER MENSUEL COMPREND-IL LE STATIONNEMENT?

Oui 3 ○ Non 4 ○ *passez à 20*

19 QUEL GENRE DE STATIONNEMENT EST COMPRIS DANS LE LOYER MENSUEL TOTAL?

		Nombre de places
Garage fermé ou stationnement intérieur.	5 ○	☐
Stationnement extérieur avec prise de courant	6 ○	☐
Stationnement extérieur sans prise de courant	7 ○	☐

Pour les codes 3, 4 ou 5 au poste 8 sur la F03, inscrivez le nombre de places (Exclure les allées)

S'IL S'AGIT D'UNE INTERVIEW SUBSÉQUENTE, MÊME MÉNAGE SEULEMENT *Autrement, passez à 21*

20 DEPUIS LE MOIS DERNIER Y A-T-IL DES CHANGEMENTS DANS LES SERVICES, APPAREILS, AMEUBLEMENT, OU AUTRES COMMODITÉS COMPRIS DANS LE LOYER?

Oui 1 ○ Non 2 ○ *passez à 22*

21 LESQUELLES DES COMMODITÉS SUIVANTES SONT COMPRISES DANS LE LOYER MENSUEL?
Cochez tous les cercles pertinents

Chauffage	01 ○	Cuisinière	07 ○
Chauffage de l'eau	02 ○	Machine à laver	08 ○
Eau froide.	03 ○	Sécheuse	09 ○
Électricité	04 ○	Autres gros appareils ménagers	10 ○
Câblodistribution	05 ○	Ameublement	11 ○
Réfrigérateur	06 ○	Aucun	12 ○

Précisez dans les NOTES

22 ÉTAT DU QUESTIONNAIRE

Rempli 1 ○ Partiellement rempli 2 ○ Refus 3 ○ *Expliquez dans les NOTES*

NOTES

N° de poste 99 ☐ N° de poste 99 ☐

99 ☐ 99 ☐

7-5000-474 06-08-89 SQC/PRI-415-62053 SQC/F-PU-015

Canada